REPORT RESUMES

ED 003 504 56

REPORT OF RESEARCH ON THE INTEGRATION OF SCIENCE TEACHING BY TELEVISION INTO THE ELEMENTARY SCHOOL PROGRAM.

BY- GARRY, RALPH J. AND OTHERS
BOSTON UNIV., MASS., SCH. OF EDUCATION

REPORT NUMBER NDEA-VIIA-031 PUB DATE 31 OCT 60
GRANT OEG-719008-09
EDRS PRICE MF-$0.18 HC-$3.48 87P.

DESCRIPTORS- *TELEVISED INSTRUCTION, *TELEVISION RESEARCH, *GRADE 5, *SCIENCE COURSES, *ACHIEVEMENT GAINS, ATTITUDES, INTERESTS, EDUCATIONAL TELEVISION, STUDENT INTERESTS, STUDENT ATTITUDES, NATURAL SCIENCES, PROGRAM EVALUATION, BOSTON, MASSACHUSETTS

DURING 1959, BOSTON UNIVERSITY CARRIED OUT A RESEARCH PROJECT TO STUDY THE EFFECTS OF TELEVISED INSTRUCTION WITH 90 FIFTH-GRADE CLASSES PARTICIPATING IN A SERIES OF TELEVISED NATURAL SCIENCE PROGRAMS. OBJECTIVES OF THE STUDY WERE TO DETERMINE HOW MUCH CHILDREN LEARNED FROM THE PROGRAMS, THE EFFECTS OF THE SERIES ON CHILDREN'S INTERESTS, ATTITUDES, AND REASONING ABILITY, AND WHICH OF CERTAIN COMBINATIONS OF AUXILIARY CLASSROOM PROCEDURES WERE THE MORE EFFECTIVE. IN ADDITION, THE STUDY WAS CONCERNED WITH THE QUESTIONS REGARDING THE RELATIONSHIP OF CHANGE TO DIFFERENCES IN SEX, INTELLIGENCE, AND PARENTAL OCCUPATION. OF THE 90 CLASSES IN THE STUDY, 18 WERE DESIGNATED AS CONTROL CLASSES AND 72 AS EXPERIMENTAL. THE 72 TEACHERS OF THESE CLASSES WERE DIVIDED INTO 3 TEACHER GROUPS--(1) THE FIRST GROUP OF 24 TEACHERS RECEIVED SCIENCE WORKSHOP TRAINING, (2) A SECOND GROUP OF 24 PARTICIPATED IN A WORKSHOP CONCERNING TELEVISION INSTRUCTION, AND (3) THE THIRD GROUP OF 24 TEACHERS RECEIVED NO SPECIAL TRAINING. TWO MAJOR SETS OF COMPARISONS WERE MADE, THE FIRST BETWEEN THE CONTROL AND EXPERIMENTAL GROUPS, THE SECOND BETWEEN AND WITHIN EXPERIMENTAL ASSIGNMENTS. MAJOR CHANGES WERE OBSERVED IN GAINS IN FACTUAL INFORMATION. OBSERVED EFFECTS ON INTERESTS, ATTITUDES, AND REASONING WERE NEGLIGIBLE. THE VALIDITY OF THE MEASURING INSTRUMENTS WAS QUESTIONED. (JC)

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

REPORT OF RESEARCH ON THE INTEGRATION OF SCIENCE TEACHING BY TELEVISION INTO THE ELEMENTARY SCHOOL PROGRAM

This research was conducted pursuant to a grant from the United States Office of Education, Department of Health, Education and Welfare, under provisions of Title VII of the National Defense Education Act of 1958.

Grant Number: 719008.09

October 31, 1960

BOSTON UNIVERSITY, BOSTON, MASSACHUSETTS

ED 003 504

PROJECT DIRECTOR: Ralph J. Garry, Ph. D.
Professor of Education
Boston University School of Education

RESEARCH ASSOCIATES: Homer J. Dietmeier
Mary Kraft
A. Cornelia Sheehan

Report Prepared by: A. Cornelia Sheehan

TABLE OF CONTENTS

PAGE

I. INTRODUCTION 1

- Problem 1
- Objectives 1
- Scope of the Study 2

II. REVIEW OF RELATED RESEARCH 3

- Television in Instruction 3
  - Science Instruction by Television 4
- Cognitive and Affective Outcomes 5
  - Images of the Scientist 5
  - Perception, Attitudes and Rigidity as Behavioral Determinants 6
- The Scientific Approach to Problem Solving 8
  - The Nature and Development of Concepts 8
  - Research Tools for the Study of Concepts 10
  - Scientific Reasoning 12

III. PROCEDURE 16

- Materials 16
  - *Exploring Nature* Series 16
- Population 20
- Teacher Assignments 21
  - Natural Science Workshop 21
  - Television Workshop 22
- Other Variable Experimental Conditions 22
  - Teaching Method 22
  - Presentation of Programs 23
  - Pupil Activity 23
- Experimental Design 24
- Testing Schedule 26
  - Administration of Pretests 26
  - Second Testing 27
  - Visits to Classrooms 27
  - Program Schedule 27
  - Administration of Posttests 27
  - Treatment of the Data 27

PAGE

IV. DATA GATHERED AND RESULTS 28

Achievement 28
  Test Construction 28
  Science Information Test 28
  Science Vocabulary Test 29
  Reading Level of Instruments 30
  Preliminary Administration 30
  Validity 30
  Reliability 30
  Item Analysis 31
  Results 32
  Experimental Teacher Population 32
  Second Testing 33
  Posttest Data 33
  Classroom Instruction and Evaluation 37
  Time Distribution 37
  Discussion 38

Science Reasoning 40
  Test Construction 40
  Preliminary Administration 42
  Reliability 44
  Results 45
  Discussion 49

Interests and Attitudes 51
  Test Construction 51
  Interest Inventory 51
  Activity Check List 51
  TV Check List 52
  Attitude Scale 52
  Function of the Attitude Scales and Rationale for their Construction 54
  Results 56
  Posttest Data 57
  Discussion 64

V. SUMMARY AND CONCLUSIONS 68

LIST OF TABLES

TABLE PAGE

1. Analysis of Variance of Science Information Scores Between Control and Experimental Groups Using Total and Partial Means by Class for the First Test Administration. 32
2. Analysis of Variances of Science Information and Vocabulary Achievement Scores from First to Third Testing Between Experimental and Control Groups Using Mean Gains by Class. 34
3. Four-Way Analysis of Variance of Achievement Gains by Mean Difference Scores for Experimental Classes, First to Second and First to Third Testing Periods. 35
4. Analysis of Variances of Achievement Gains from First to Third Testing Using Means by I. Q. Quartiles for a Random Sample of 500 Subjects Drawn from the Experimental Group. 36
5. Means of Information and Vocabulary Achievement Scores for First and Third Testing Periods by I. Q. Quartiles.. 37
6. Correlations Between Instruments Used to Study Science Achievement. 39
7. Means, Standard Deviations and Reliabilities of the Three Science Reasoning Tests. 44
8. Analysis of Variances of Pretest Data on Science Reasoning Tests Between Control and Experimental Groups. 45
9. Analysis of Variances of Pretest Data on Science Reasoning Test II Among Experimental Classes and of I. Q. Data Between Experimental and Control Classes. 46
10. Means and Standard Deviations for I. Q. Data and for Science Reasoning Test I Gains, Between Experimental and Control Groups by Total Class and by Sex. 46
11. Analysis of Variances of Gains in Science Reasoning Test I Between Experimental and Control Groups by Total Class and by Sex. 47

TABLE PAGE

12. Four-Way Analysis of Variance of Science Reasoning Gains by Mean Difference Scores for Experimental Classes, First to Third Testing Periods. 48

13. Correlations of Instruments Used to Measure Science Reasoning with Otis Beta I. Q., Using Pretest Scores and Difference in Gains Scores. 49

14. Correlations Between Instruments Used to Measure Science Reasoning. 50

15. Mean Scores by I. Q. Quartiles for First and Third Testing Periods on Instruments to Measure Interests, Activities and Perception of Scientists. .. 57

16. F-Values for Analysis of Variances by I. Q. Quartiles Over Testing Periods for Interest, Activities, and Perception of Scientists. 58

17. Four-Way Analysis of Variance of Changes in Interest, Activities, and Attitude (Part II-B: Perception of Scientists) by Mean Difference Scores for Experimental Classes First to Second and First to Third Testing Periods. 59

18. Relationship of High and Low Science Interest to Science-Related Television Viewing Employing Phi Coefficients Converted to Tetrachoric Correlations. .. 60

19. Partial and Multiple Correlations of Critical Variables of Perception of Scientists and Attitude Toward Science and Scientists with Six Predictor Variables Related to Children's Values and Beliefs. 61

20. Analyses of Variances of Changes in Perception of Scientists and of Consistency of Beliefs and Values (Attitude) by Status of Subjects on Predictor Variables Derived from Attitude Scale Parts I-A and II-B. 61

21. Correlations Between Instruments Used to Measure Interests and Attitudes. 62

22. Correlations Between Predictor and Criterion Variables Used to Measure Changes in Attitude Toward and Perception of Science and Scientists. 63

I. INTRODUCTION

Problem

The research on the use of television as a teaching tool points almost consistently to the conclusion that television as an instrument of instruction can be at least as effective as, and under certain conditions, somewhat more effective than, conventional modes of instruction for acquisition of factual knowledge and mastery of some skills. Two problems emerge from the current status of research on this medium (1) the results suggest that television's effectiveness as an instrument of instruction is contingent upon certain requisite conditions of use and that degrees of effectiveness are related to these conditions and combinations of conditions. Indiscriminate and unstudied use of television as a teaching medium gives promise of being at best inefficient and wasteful and at worst ineffectual or even detrimental to learning. The implication here is that the implementation of educational television, and not the basic usefulness of the medium, is becoming the research issue, and it is treated as such herein; and (2) the traditional interpretation of effectiveness has been in terms of educational achievement--facts, knowledge, information acquired. The research endeavor to be recounted here is concerned with two additional interpretations of effectiveness:

1. There has been a lack of study to date of the possibility that televised instruction, particularly in the sciences, will increase the subject's ability to do scientific reasoning and develop an ability to use the scientific method to solve problems. Since we are living in a scientifically oriented culture, the feasibility and the effectiveness of the television medium in the development of children's understanding of the scientific method and the place that it holds in our way of living and thinking would seem to be of justifiable research as well as educational concern.

2. While facts, knowledge, and information may be transitory and easily dissipated by the interference of new facts, knowledge, and information, there is another dimension of organization within the organism which is perhaps of even greater concern because of its relative stability and canalizing function. Effectiveness, then, can as well be defined here as the creation or intensification of a favorable attitude toward science and scientists and evidence of increased interest in scientific activities.

Objectives

The project constituted a study of the effects of teacher training, method of presentation of program, teaching method and pupil activity upon the achievement, interests, attitudes, and scientific reasoning ability of fifth grade pupils studying natural science by television. With regard to science achievement, interest in science, attitude toward science and scientists, and science reasoning ability of fifth grade pupils:

1. Are differences observed as a function of specialized training of teachers in (a) a workshop in science for the elementary school or (b) a workshop in the utilization of television in teaching?

2. Are differences observed as a function of using the televised natural science programs to (a) initiate learning in connection with a topic or unit or (b) terminate learning in connection with a topic or unit?

3. Are differences observed when (a) teachers use a study guide prepared for use in conjunction with the programs or (b) teachers direct the learning independently of the study guide?

4. Are differences observed when (a) pupils are given a common class assignment in connection with the programs or (b) pupils are encouraged to develop individual interests and specialties through individual projects in connection with the programs?

5. Are differences related to other variables such as (a) intelligence, (b) sex, (c) teacher interest in science, or (d) parental occupation?

6. Are differences related to attitudes held by the pupils prior to their exposure to the televised natural science series and correlated class activities?

Scope of the study

The study being reported here was made possible by a grant from the United States Office of Education of the Department of Health, Education and Welfare, through the New Educational Media Branch under the National Defense Education Act of 1958. The research utilized approximately 2600 children in 90 fifth grade classes in 23 towns and 7 cities randomly chosen from all available fifth grade classes in some 30 participating communities within a fifty-mile radius of Boston, Massachusetts. The study required the training of 24 of the ninety teachers in the utilization of educational television in the elementary curriculum; the training of another 24 teachers in the implementation of natural science instruction in the elementary grades; the provision of "no special training" for another 24 teachers; and designation of the remaining 18 teachers as control teachers. Testing was conducted in the fall, the winter and the spring in conjunction with the reception by the ninety classes of a series of thirty televised natural science programs during the period October, 1959 through May, 1960.

II. REVIEW OF RELATED RESEARCH

Television in Instruction

Historical perspective. Televised instruction in the classroom was initiated more than two decades ago with a telecast 45-minute presentation to 250 science students at New York University (Clark, 20). Experimental productions carried on in colleges and universities from that time, except for the interval of the war years, bear witness to the recognition of the potential of television for education. Notable among those spearheading the instructional use of television was Iowa State College, among the first to own and operate its own television station and provide courses at the elementary and secondary levels as well as at the college level (Dunham and Lowdermilk, 30), Brooklyn College, Syracuse University and, using television in the public school system, Philadelphia, New York, Baltimore and Pittsburgh (Dunham and Lowdermilk, 30; Gable, 36; Gross, 40; Brooker, 10; Steele, 84).

During this exploratory period there was little in the way of formal evaluation of television teaching as compared with conventional instruction. Initial research in the instructional use of television was conducted by the U. S. armed forces (Desiderato, 25; Kanner, 49; and Kanner, et al, 50, 51) and from the results of these studies it can generally be concluded that televised instruction is at least as effective as regular instruction, on occasion more effective with lower-aptitude groups, and retained at least as well as regular instruction. From extensive experimentation and evaluation, the armed forces have demonstrated the flexibility of television in meeting training requirements by increasing class size, securing better utilization of good instructors, and showing close-ups of small objects.

In 1954 under the auspices of the Ford Foundation and the Fund for the Advancement of Education, Pennsylvania State University entered a comprehensive program for research in the use and evaluation of instructional television in higher education (Carpenter and Greenhill, 17; 18). Utilizing closed-circuit television in a wide range of college courses, television was shown to be as effective as personal instruction in every instance. Retention studies showed the materials learned by television were retained as well as those learned through conventional means of instruction.

Research on the use of television in schools has been less abundant and less systematic than either the college or military endeavors. Problems have been encountered at all levels, from the modern philosophy of educational practice to those dealing with scheduling and physical facilities, and the heterogeneity of instructional groups. Much of the in-school television to date has been of an enrichment nature, supplementary to and not directly integrated with regular classroom instruction. An exception to this was the experiment in direct teaching by television conducted by the St. Louis public school system. Herminghaus (43) reports that the results, while showing that in many cases the experimental classes did equally as well as the control classes, pointed to the fact that television for a variety of reasons could not provide total teaching, but needed some follow-up activity by classroom teacher for maximum effectiveness.

By its very magnitude alone, no other single experiment in the educational use of television has drawn so much attention as the Hagerstown Project, Washington County, Maryland, launched in 1956 to continue for a five-year period. Using closed-circuit television, lessons covering a broad range of subjects reach many schools with thousands of pupils from primary grades through high school. Buehring (12) reports that, although conservative in making premature claims of results, the Hagerstown project points tentatively to the equality and even the superiority of televised instruction as compared with regular classroom instruction. Consistently, evaluation of in-school television has taken the form of achievement testing.

Hoban (45) has compared the present status of television research with that of motion pictures and radio when so much effort, largely wasteful, was expended on comparative effectiveness studies during their development. The basic question for research today is not whether instructional television is equal to, less or greater than conventional teaching in effectiveness. The question should rather be for what are television and conventional instruction best suited respectively, and how can they be combined to produce optimum results. The problem of defining specific teaching-learning functions and of determining how television may relate to them has not yet been approached systematically. Researchers need now to turn to exploration of television utilization for itself to try to discover the factors which operate with television to promote more effective teaching-learning situations. Clearly, television has significant potential for contributing to the learning process at many points, but only when it is properly integrated with the teaching purpose and the learning need in the individual classroom (Levenson, 57; Wigren, 90). The need for clarification of the classroom teacher's role in the utilization of instructional television is emphasized by many of the more recent studies (Seipmann, 82; Klapper, 53).

Science instruction by television. (54) Program logs from local communities show science to be the subject most frequently included in school television programs. Satisfying experiences with science via television through home viewing have already favorably disposed children to the medium and the materials in this area (Nelson, 70; Campbell, 16; Yanowski, 93). The audial and visual attributes of television are stimulating forces for science learning. Classroom teachers generally agree that science telecasts are superior to all others (Campbell, 16; Yanowski, 93). Television appears to be particularly suited to biology instruction where lack of adequate equipment and experimental materials often limit good teaching in this area (Draper, 29; Reiner, 75). Television provides each student with a powerful microscope. Live animals and plants of a wide variety can be used in demonstration or viewed in their natural habitats. Today children are engaged in school-originated and leisure-time science projects at an unprecedented rate. In this age of science, it is conceivable that television, if properly used, will improve the quality of instruction, bring real life to science teaching, extend master instruction to more schools and motivate young people to consider careers in science.

Cognitive and affective outcomes

Practically without exception, the data reported from the various experimental uses of instructional television around the country consist of results compiled on the basis of subject-matter tests. The learning of subject-matter is, of course, only one of the objectives of an educational program. Other outcomes do not lend themselves as readily to paper and pencil test, however. Development of attitudes, problem-solving ability, human relations, improvement of study skills, all are not measured by the tests used by the experimental projects reporting to date. Whether pupils who received instruction by television did as well in these areas as in achievement is not as yet determined. Nevertheless, a 1957 seminar of the National Association of Educational Broadcasters (86) concluded with the recommendation that research involving affective behavior, critical thinking, and problem solving aspects of cognitive behavior probably would make the greatest contribution at this time to understanding and more effective use of the educational mass media. A report from a National Education Association seminar (85) on the role of television in instruction (1958) lists some of the ways in which television could make a contribution to better education, among them, providing motivation and stimulation, developing attitudes, stimulating creative activity, promoting problem solving and critical thinking, and challenging pupils to assume more responsibility for their own learning. If, in fact, the foregoing are conceived of as worthwhile objectives of televised instruction, steps should be taken to devise techniques whereby their realization can be evaluated.

Images of the scientist. Polling a nation-wide sample of high school students on their ideas and feelings about science and scientists, Remmers and Radler (77) found that 25 per cent of the students think scientists are willing to sacrifice the welfare of others to further their own interests, 28 per cent believe that scientists don't have time to enjoy life, 30 per cent declare that one can't raise a normal family and become a scientist, and 35 per cent believe that it is necessary to be a genius to be a good scientist.

In a large-scale study with similar objectives, Mead and Metraux (67) employed sentence-completion essays written by high school students to determine adolescent attitudes toward and perceptions of science and scientists. They found that the adolescent image of the scientist, without personal career involvement,was highly favorable (and largely correct); but as a choice for personal career, or choice for husband's career, the image was a very negative, unfavorable and largely distorted one. Mead and Metraux (67) draw implications as to the roles of the classroom teacher and the mass media in the development of student attitudes toward science and challenge each to assume its responsibility in the realistic portrayal of science and scientists.

Reviewing significant research on the influences of television on the American culture, McPhee (66) concludes that mass communication appears to canalize existing predispositions rather than change them---people expose themselves to programs in line with their existing predispositions. Thus the mass media would seem to hold more promise in the reinforcement of present knowledge and attitudes than in their modification. Nevertheless, there is reason to believe that a face-to-face follow-up of the exposure to mass communication can enhance its potential for selling ideas or modifying attitudes or behavior. If self-selection is a problem of home viewing (in the sense of compounding the restrictive effects of predispositions), consideration should be given to the possibility that there may be a selective-perception counterpart in the classroom viewer.

Perception, attitudes, and rigidity as behavior determinants (81). Bills (7) notes that behavior is consistent with the individual's perception of the world in which he is behaving. The individual's perceptual field, however, is a function of many component factors including needs, values, beliefs, attitudes, interests, experiences, etc. The inability to maintain a flexible or fluid perceptual field may be due to a lack of experience or opportunity to make new discriminations or it may be due to the need to defend and maintain the present perceptual organization. The more rigid the perceptual structure, the less reality-oriented will be the responses to new stimuli. Bills (7:26) states further, "The person whose perceptual field is labile and flexible, and who thus is able to make the unique response required by the newness of any situation, may be described as open to experience.....his present perceptions are not a distortion of his past experience.....his experience is immediately available to him."

Murphy (69), also recognizing the selective and distorting function of perception, states that anticipation and expectation based on prior experience give meaning to each new experience. "This means that not only recall and thought but the very process of perception is itself warped from the beginning by prior experience. One sees and hears in terms of one's expectations--expectations that are almost never free of affective coloring." (69:55) On the subject of the influence of attitudes on behavior, Remmers (76:7) notes that "the realization is rapidly growing that attitudes, the way individuals and groups feel about the various aspects of their world, are probably more determinative of behavior than mere cognitive understanding of this world. When this is granted, the importance and value of attitude measurement becomes at once obvious."

Peak and Morrison (73) offer experimental evidence that while selectivity resulting from attitudinal position may not necessarily influence what is remembered it does affect the assimilation of information into an existing attitude structure. An individual may hear and remember a piece of information which is contrary to his beliefs (as manifestations of an underlying attitude), but the communication may remain merely as a recollection that a certain statement was made rather than being accepted as relevant to a personal position on the subject, unless the statement was to some degree consonant with an existing attitude. Peak and Morrison also report results (73:132) which suggest differential sensitivity to communications of subjects at the two extremes of the authoritarianism continuum as measured by the California F Scale (1). The narrowing of perceptual sensitivity (increased selectivity) with correspondingly increased specificity of response to stimuli perceived as supportive to and congruent with an attitude and an increased tendency to reject stimuli which are incompatible with and/or threatening to the prevailing attitude structure can be characterized as a rigidity syndrome. In the case of the child, the more nearly the attitude held is consistent with the values of the adult culture and the more nearly the child's beliefs and convictions, as exemplary of the attitude, coincide with those of significant adults and with peers, presumably the more resistant to change will the attitude be. Because this narrowing of the perceptual field appears to be a cumulative function of the dynamic interaction of personality and social influences under specifiable circumstances, susceptibility to change may decrease with time as a function of the individual's experiences and his perception of his status in the social matrix.

Murphy (69) proposes that because the object of self-reference is always evolved within a family and a community group through identification with significant figures, it is impossible to imagine an instance where the importance of the group is not consciously or unconsciously, directly or indirectly, overexaggerated. "It is characteristic of every social group to develop its own socially shared forms of 'loaded' perceiving and thinking....built up through a great deal of consistent social sharing." (69:120)

Identifying five characteristics of attitudes: direction (for-against); degree (favorable-unfavorable); intensity (degree of confidence with which stated); conscious-unconscious; and conflicting or inconsistent; and using a Likert-type scale (59) calling for responses of *true, false* and *don't know* (beliefs) and *approve, disapprove*, and *don't care* (attitudes), Garber (37) found that the following relationships held:

1. When a statement is believed true, retention of the statement is enhanced.
2. Approval tends to enhance retention and disapproval to inhibit retention, but not to a statistically significant level.
3. When belief and attitude reinforce each other positively, there is positive facilitation of retention.
4. When belief and attitude reinforce each other negatively, there is inhibition of retention.
5. When belief and attitude are conflicting or inconsistent, the attitude dominates over the belief in influencing behavior. (37:389)

Cartwright (19) in an exposition of contemporary research in the study of social-psychological processes notes that a hypothesis which is receiving much conceptual attention and apparently considerable methodological confirmation is that when "group anchored" attitudes are exposed to counterpressures their resistance to change will be greater when the group has higher salience for the individual. With the movement away from descriptive research and toward the study of functional relationships among constructs, Cartwright points out, there arises the need for a defensible rationale to govern the way in which such research findings are to be generalized.

There exists in the literature an apparent controversy as to whether behavioral rigidity is a situational phenomenon or a generalized personality characteristic. Ainsworth (2) and Levitt (58) both reviewed the research in this area and concurred that there is no definitive research on the subject of rigidity to date. Luchins (60)--using the water-jar Einstellung test--maintains that situational factors determine rigidity of performance of problem solving; on the other hand, Rokeach (78) and others have tried to show that rigid behavior in the problem-solving situation can be ascribed to a generalized personality trait which precipitates rigidity of attitudes and behavior in a wide range of situations. Ainsworth has evolved the following conceptual position in the belief that it offers possibilities for reconciling the two positions: (2:67-8) (1) rigidity is attributable to insecurity, (2) insecurity may be induced by the stressful properties of the situation but also by the rearousing of feelings associated with other insecurity-producing situations, (3) situations hold differential degrees of stress for different individuals depending on the degree to which the subject perceives himself as adequate to the situation or the degree to which he generalizes his insecurity from previous experience, (4) rigidity emerges as a defense against insecurity in a particular situation and is tied to generalized rigidity only through the intervening variable of insecurity. Thus the so-called rigid or defensive individual will manifest rigidity in problem-solving only to the degree the situation provokes insecurity. Employing the Einstellung technique, Ainsworth found experimental support for his hypotheses derived from the above position. He suggests further that "insecurity in problem-solving situations manifests itself in tendencies toward a diffuse and disorganized response, and it is against these tendencies that rigidity acts as a defense." (2:74)

Meresko et al (68) offer the following conceptual formulation of psychological rigidity: "a person's resistance, or lack of readiness, to be influenced by motivationally relevant stimulation in such a way as to adjust to his environment as effectively as his behavior-repertory permits....(including) both defensive inhibition (resistance) and organic deficit (lack of readiness); both reactions to external stimulation and reactions to inner processes; both rigid cognitive functioning and/or rigid action." (68:89) These authors feel that this widely conceived position is more useful than some of the more limited schema proposed to handle rigidity:

> "It embraces 'inappropriate perseveration of a set' as 'resistance, or lack of readiness, to be influenced by a change in the situation.' It also embraces 'resistance, or lack of readiness to be influenced' by relevant features of one situation considered by itself: 'intolerance of ambiguity' and 'narrow-minded' concept formation based on resistance to surveying freely a whole that contains threatening features, cognitive 'flight' from the crux of a situation into focus on isolated details, or into vagueness, formalistic overgeneralizations, or mysticism, and so on." (68:89)

The Scientific Approach to Problem Solving

The nature and development of concepts. Curti (24:177), surveying the research on concept development, defines concept as "a present item of experience functioning in a symbolic capacity as a response to any one of a class of objects or relationships having a common characteristic." Her failure to clarify the meaning of "item" gives unfortunate latitude to those who would correlate words with concepts. At another point in her review, Curti proposes a number of characteristics that a concept must have, among them: (1) a concept is built up of repeated experiences, (2) concept development is a gradual process, and (3) logical and flexible concepts cannot be acquired through mere word learning." Leeper (56) reporting on cognitive processes, uses "concept" and "cognition" interchangeably, although implicit in his usage is the conclusion that cognition is the application of concepts to problem solving. Furthermore, Leeper (56:748) lists six factors which interfere with the formation of new concepts:

> "1. A separation of time between cause and effect (as between mosquito bites and malaria).
> 2. Perceptual incongruities between the cause and effect (as the difference between flies and maggots, in the old theory of 'spontaneous generation').
> 3. Already existing habits that cause a misleading organization of the items in the situation (as the knowledge that led early workers on beri-beri to search persistently for the transmission of bacteria from one affected person to another).
> 4. Relative intangibility of either the cause or the effect, or both (as the elusiveness of wind resistance as a factor in the inefficiency of trains at high speed).
> 5. The existence of interdependent causes (as the fact that neither sufficient calcium nor sufficient vitamin D can prevent rickets by itself, but that both must be present).
> 6. Motivational factors that work against acceptance of a conclusion (as when Semnelweiss's fellow obstetricians in the 1840's refused to accept his demonstration that puerperal or child bed fever was due to the transmission of some material from one woman to another by the obstetricians themselves)."

One of the most complete analyses of concept formation and perhaps the most meaningful is that of Vinacke (88). Acknowledging the complex nature of the process under study, Vinacke points out how the study of concepts is further complicated by the complex interrelationships which exist between concept formation and the nature and development of language, perception and social interaction, plus by the necessity to work primarily by inference since so little is known about the internal processes of the nervous system in this respect. With consideration for the complexity of the problem, Vinacke suggests the parameters of research in this area are as follows:

1. We must clarify what we mean by the process of concept formation.
2. We must differentiate between the process of concept formation and the concepts themselves.
3. We must differentiate the genetic aspects of concept formation from the problem solving aspects of concepts.
4. We must differentiate between concept formation as it is found in children and as it is found in adults.

Another outstanding contribution to the understanding of concepts has been made by Brownell (11). His approach is founded on the importance of meaning and he proposes that all learning can be placed somewhere on a continuum of meaningfulness, depending on the individual learner. Since, Brownell maintains, concepts are dependent upon meaning, concepts should also lie somewhere on this same continuum. The value of Brownell's idea can be further extended if one separates "completeness" of a concept from "meaningfulness" of a concept, and constructs a dual continuum to handle these interrelated though separate attributes of concepts. This dual continuum permits us to perceive how a child can be taught a relatively complete concept and yet have learned a concept that has little or no meaning.

Research tools for the study of concepts. The interview or question-asking technique was developed by the Swiss researcher in child development, Jean Piaget. This technique and Piaget's experiments have been repeated or extended by other investigators, one of the most recent of which is the study by Deutsche (27) of 732 children ranging in age from 8 to 16 years and covering a wide range of socio-economic groups. Deutsche used a series of demonstrations calling for some kind of explanation, for instance, placing a glass over a burning candle and leaving it until the flame goes out. While this technique has value in that it deals with concepts as they are actually found in a life situation, considerable care must be taken to insure that the response is a real concept understood by the child and not the repetition of a ready-made concept given to the child by parent or teacher. This is a difficult discrimination to make, since the child who really understands may choose to express his understanding in the words of the adult, and likewise the child who has merely memorized a ready-made concept may be quite facile in the use of words and be expressing something he doesn't fully understand.

Other investigators have been searching for a testing technique possessing two major characteristics: (1) it should be a test in which a "new concept" would develop during the testing, and (2) it should be a test that could be administered to adults as well as children. Although not originally designed for the study of concepts, the Vigotsky block test, later redesigned by Hanfmann and Kasanin (41), has become a popular technique for the study of concepts which meets the above qualifications. The blocks, 22 in number, are classifiable in four groups based on a combination of height and area with color and shape operating as distractors. The process involved in the correct sorting of the Vigotsky blocks is similar to the problem facing the scientist who is seeking an unknown cause and effect relationship--there is absolutely no one who can help him to find the unknown factors and he knows only when he has failed. While this is the way that a scientist works, it is not the way in which children learn concepts which have already been discovered by adults. We should be teaching children how to search for new concepts but to make such a method the sole method of teaching concepts would make education practically an impossibility. Another limit of the Vigotsky block test is that the relationships between the elements is a static, unchanging one and the concept is strictly one of classification. Most of the life experiences of human beings are involved with dynamic relationships and processes and dynamic interaction brings with it the need for predictability. How much predicting can one do with a classification concept? The classification concept becomes of value only when it becomes part of a dynamic process and even then the classification concept has questionable value in situations involving human interaction. While we need to study classification concepts, it would seem that we have a more urgent need to study process concepts.

Another block test devised by Maag (61) is subject to practically the same limitations as the Vigotsky blocks, but is even better for observing the scientific method of investigation because of step-by-step features for observation of subject actions. Classification tests have also been constructed around card sorting. Heidbreder (42) used 144 cards containing a total of nine concepts. The sorting process permits the observation and quantification of such overt actions as groupings made, changes made in groupings, correct and incorrect sortings, periods of hesitation, verbal comments, etc. The main limitation of Heidbreder's cards lies in their extreme simplicity. The images used on the cards were originally used for a memory test and involve relatively simple concepts. With the elimination of the memory factor, many of the catergories become overly obvious. This limitation, however, could be overcome by designing pictures specifically for this purpose. The picture feature of cards offers many possibilities since the pictures can be related more nearly to life situations. It might be possible to have each picture-card represent an element of a total concept and even to do this in such a way that the test could deal with concepts of process instead of the present limited concepts of classification.

A number of investigators have built concept tests around standard and commonly used words, among them, Reed (74) and Underwood and Richardson (87). While it may be possible to construct a conceptual test using words, both of these are classification tests; moreover, since they use nothing but words as the element of the concept, the test does not involve sensory experience as such. Actually further analysis reveals that both tests rely on visual memory. This is particularly the case with the Underwood-Richardson test, where, for instance, one must resort to the mental image of a barrel as "round" because no visual image of this sort is provided in the testing process. While one could legitimately construct a test using the principle of mental images, in so doing, one would be testing concepts already formed rather than testing the development of new concepts. In short, concept tests that involve words only tend to measure past concepts and have the tendency to overemphasize classification.

Because of the fact that concepts and concept formation seem to be related in some way to the process of perception, a number of tests have been constructed using abstract visual designs as the primary stimulus. Detambel (26), Newman (71) and Kurtz (55) all designed tests of this nature, and the tests have obvious similarities. Each uses two elements to form a concept, each is merely a test of classification or recognition, and each depends upon a verbal response of some nature in order to solve the problem. All three tests depend on an interrelationship between two separate parts of a design. In Detambel's test it is the position and the height of a rectangle; in Newman's test it is the azimuth of a line and the position of a cluster of dots; in Kurtz's test it is the nature of a figure (circle, square, small, large) and the position of the figure.

While the use of a single key element is much simpler than the use of two, simplicity is not the only reason for not using a single element. If we consider, for instance, "rain" as a concept, falling drops of water can be understood in two basically different ways. In one it is nothing more than an experience, visual and tactual in nature. In the other it is a more complex concept involving hydrodynamics and gravitational laws. One does not have to understand the second in order to be able to recognize the first. This is why an abstract design concept test cannot use a single element--when a single element is used, it becomes a test of recognition not a test of conceptual learning.

There seems to be considerable difference between concepts as defined theoretically and concepts as they are found operationally defined in research designs. The experimentation has tended to emphasize the classification aspect of concepts and completely ignore the more dynamic features of concept formation. For this reason the interpretation of most research on the subject must be limited to concepts of classification and no generalizations can rightfully be made from these experiments with regard to concepts of process. This is most regrettable, for the pressing needs in the teaching of concepts lie in the teaching of concepts of process. The direction that research designs should be taking, therefore, is toward studies which will more closely relate to the more complex and dynamic aspects of concept development.

Scientific Reasoning. Before undertaking research on scientific reasoning (28) it is necessary to arrive at a workable definition of scientific reasoning. This is a very difficult undertaking since scientists themselves do not agree on the nature of scientific reasoning. The general opinion among scientists seems to be that it is impossible to limit scientific reasoning to any set pattern of thinking or set procedure of operation.

The problem of defining scientific reasoning has been further complicated by the use of the philosophical terms inductive and deductive. Through popular usage inductive has come to mean reasoning from the part to the whole or from the particular to the general, while deductive has to come to mean reasoning from the general to the particular. These popular definitions have done much to confuse the issue regarding scientific reasoning. Since the scientist does often start with individual facts and then work toward some general law that will include these facts, it has been popular to assume that the scientist is one who uses primarily inductive reasoning. This assumption becomes highly questionable when one asks the question, "How does the scientist check the validity of his general law?" He uses basically two methods, both of which involve deductive reasoning. In the one case he uses deductive reasoning to see if his known facts do logically follow from his general law and in the second case he uses deductive reasoning to arrive at a logically derived fact which he must then test experimentally. This problem of inductive reasoning vs deductive reasoning has been discussed in some detail by Bayles (5) who argues quite convincingly that the modern scientist uses both inductive and deductive reasoning.

Several individuals, Crowell (22), Keeslar (52), Burmester (14), have approached the problem by compiling lists of attributes and skills involved in the scientific method of thinking. Keeslar maintains that educational research in the teaching of science is being held back by the reluctance to define in detail the steps involved in the scientific method and the confusion of scientific method with scientific attitude. Keeslar compiled a list of 42 elements involved in the scientific method. After submitting his list for evaluation to 22 research scientists on the staff of the University of Michigan, he condensed the list to 10 major elements and 17 minor elements. Burmester (15), using Keeslar's work as one of her primary sources, developed a list of eight abilities involved in the scientific method:

1. Ability to recognize problems.
2. Ability to delimit a problem.
3. Ability to recognize and accumulate facts related to the solution of a problem.
4. Ability to recognize a hypothesis.
5. Ability to plan experiments to test hypothesis.
6. Ability to carry out an experiment.
7. Ability to interpret data.
8. Ability to apply generalizations to new situations.

In the research reported in this paper it was assumed that scientific reasoning is both inductive and deductive in nature and that it involves the abilities outlined by Burmester.

One of the major problems in doing research on scientific reasoning is that of testing scientific reasoning ability. Bloom (8) has a number of excellent examples of various types of questions which could be used in the construction of a test. Tests have been constructed by Baggaley (3), Dunning (31), Edwards (33), Engelhart (34), and Nelson (70). Brandwein (9) discusses in some detail the problems of constructing science tests for the college board examinations. Most of these tests involve a great deal of reading or they require individual administration. Those tests which can be scored by machine have the common characteristic of giving clues in the multiple choice answers. These clues can be used in order to make an intelligent guess as to the correct answer; consequently, the test scores measure a combination of intelligent guessing and reasoning. Tests that eliminate intelligent guessing require either written answers or verbal answers, the former cannot be machine scored and the latter requires individual administration.

Although there has been very little research with scientific reasoning in the elementary grades, there is considerable evidence that such reasoning ability does exist and can be developed. Inhelder and Piaget (47) using observation of individual children have designated what they believe to be four fairly clearly defined stages in the development of a child's ability to use reasoning in the solution of problems. Before the age of 7-8 the child is concerned with the success or failure in solving the problem and not much concerned with reasoning. Between the ages of 7-8 and 11-12 the child becomes more interested in the concrete operations of solution and is quite capable of pointing out the important features in a particular solution; however he shows little ability to abstract the situation. Between the ages of 11-12 and 14-15 abstraction and formalization begin to develop. Formalization does not reach its peak until after age 14-15. Unfortunately Inhelder and Piaget operate entirely on an observational level and seem to assume that the development of reasoning is a maturation process. They make no attempt to teach or develop reasoning ability. They also make no attempt to find out why a particular child may have developed faster than another, i.e. there is no data regarding the background, educational or socio-economic, of a particular child or group of children being studied.

Several individuals have observed reasoning ability in children at fairly early ages. Hill (44) noted that children in the elementary grades speculate before they make a decision. Webb (89) observed that even the 2½ year old has ability to generalize from past experience. She also noted that young children will tend to reason even though their data may be quite meager. John and Miller (48) noticed that *some* children between the ages of 6 and 12 did as well in the solving of mechanical reasoning problems as did *some* doctoral candidates who attempted to solve the same problem. West (90) and Dunning (32) have criticized the present methods of teaching science at the elementary school level. Both feel that there is not enough emphasis placed on the reasoning process. In a more formal experiment on generalization Croxton (23) found that there was an increase year by year in children's ability to make generalizations. He noted that 62% of the fifth grade children under study indicated ability to generalize. Nelson (70) found that elementary children were quite capable of understanding certain scientific concepts of light and sound.

There have been a few experiments undertaken to determine the effect of scientific training on information learning and problem solving. O'Connell (72) found that high school students who learned chemistry by an inductive method did better than those who learned by the more usual deductive method. Inductive method in her study referred to a method of teaching which started with laboratory facts and reasoned toward general laws rather than the usual method of teaching the general law first and then demonstrating it in the laboratory. Marks (63) found that formal training made no difference in the ability to solve problems. Mason (64) found that those who received direct training did better in scientific thinking than those who did not receive such training. The major difference between these two studies was that the training in Mark's study was by lecture while the training in Mason's study was carried out in the laboratory. Mason's subjects received practice in scientific method while those in Mark's study did not. In part these differences in training methods help to explain the difference in the two experimental results. Craig (21) found that the poorest method of teaching problem solving was to give the subjects the solution. He found that the next best method of teaching was to point out the errors made in attempting a solution and that the best method was one in which the student was given certain specific directions along with pointing out the errors.

There are definite indications that set is an important factor in problem solving involving reasoning. While Battig (4) found that the individual who used the most consistent approach usually achieved the best results Goodnow and Pettigrew (38) found that patterns learned in earlier solutions affect the methods used in later learning situations. If the later situation does not lend itself to the earlier approach there will be a decrease in efficiency. Youtz (94) obtained similar results when he found that a successful hypothesis in one problem would block the use of a new hypothesis in a new problem. Maltzman et al (62) found no relationship between anxiety and problem solving or neuroticism and problem solving; however, Gordon and Durea (39) found that when subjects were told during a preliminary practice period that they were doing poorly, they did not do as well in the testing situation as those who had not received such criticism.

Other factors that might be involved in reasoning were indicated by a number of experimenters. Hunter (46) noted that younger children seemed to be more affected by the testing situation, benefited less from practice and found the abstract more difficult than the concrete. Burack (13) concluded that there is no one way to approach problem solving, that the approach varies with the individual and within the individual from problem to problem. Saugstad (80) found in a study of high school students, a negative correlation between incidental memory and problem solving ability. He found further that there was a significant difference between boys and girls with the difference favoring the boys. Billings (6), who came to the conclusion that problem solving was a general factor after he discovered a high correlation in problem solving ability in different fields of specialization, also found men to be superior to women. McNemar (65) in a study using college undergraduates found a similar difference between men and women in reasoning ability.

While there have been very few studies made of the effects of audio-visual materials on reasoning ability Farrell (35) and Woodburn (92) have made strong pleas for the multisensory approach as the basis of learning. They argue that more emphasis should be placed on primary sensory information in order to preserve a better balance between reasoning from observation and reasoning from verbal symbols.

Since it is assumed that watching motion pictures is a passive situation it has been concluded by many that films could be of little benefit where reasoning is involved; however the little research evidence that is available does not support the prevailing attitude. Rulon (79) found that subjects taught by motion pictures did better on both rote and eductive (reasoning) items than students who were taught by text book.

III. PROCEDURE

Materials

The experimental materials were a series of 30 half-hour natural science programs entitled, Exploring Nature, produced by the Massachusetts Audubon Society and designed to develop understanding of the intricate relationship of all living things to their environment and covering such topics as (a) the nature of life and living things, (b) cell theory and multicellularity, (c) the flow of energy in the living cell, organism and community, (d) reproduction, heredity and development, etc. These programs, resumés of which follow, presented by the "21-Inch Classroom" of the Eastern Massachusetts Council for School Television, and telecast by WGBH-TV, Channel 2, in Boston, were viewed by all classes in the study beginning in early October, 1959, through May, 1960. The TV teacher was Professor William H. Weston, Ph. D., Harvard University faculty, Department of Science.

Exploring Nature Series

1. Alive or Not Alive. This program discusses the distinction between living and non-living matter. It is illustrated by model systems which mimic the various characteristics of life--such characteristics as the use of energy, growth, reproduction, and response to stimuli are treated.

2. These are Animals. The characteristics of animals is the central theme of this program. Movement as a means of obtaining food and the use of food as a source of energy are discussed.

3. Plants. Plants are living organisms capable of trapping the sun's energy and storing it for future use by themselves and other organisms. Plants control both the numbers and the distribution of some sorts of animals. This program is primarily concerned with the role of plants in the living world.

4. Animal Tools and Weapons. The principal use of animal tools is to secure food; their use as weapons is secondary. Some animals, however, possess structures that are used solely for defense. The variety of animal tools and weapons is enormous but all are of value--each is selected to do its job well.

5. Mammals. The various characteristics of mammals and the advantages that these characteristics confer on their possessors is the central theme of this program.

6. How Animals Move. Animals move in various ways. The basis of this movement, however, is the same in almost all animals--the contraction of a muscle against a fixed object. This program illustrates the action of muscle against bone and compares relative advantages to vertebrates and invertebrates.

7. Animal Homes. Homes serve for the protection of the young, for survival of the adult and for the storage of food. The nature of the home of the animal that lives in it, and the variety of homes, are almost as diversified as the animal inhabitants. This program emphasizes how homes are used for the protection of the young and why they are necessary.

8. Seeds. The nature of seeds, the variety of methods by which they are distributed, their role in the life of the plant and the process by which they germinate, are discussed and illustrated in this program.

9. Trees. Trees are rather specialized plants. The methods by which they grow, obtain raw materials to make food, and by which they survive the winter are shown. The economic importance of wood and of forest practices are included.

10. Water. All life depends upon water. Its role in the weathering of rocks, the origin of life and the sustaining of life, are considered. That all animals and plants still live in a watery environment as an aggregation of cells is made evident.

11. Fossils. Fossils are records in the rocks. They are found in various sedimentary rocks and from them we can learn about life that was present long ago. One of the greatest generalizations in biology, the theory of evolution, is largely based on a knowledge of the fossil record. The emphasis in this program is on the role of fossils in the theory of evolution. Formation of fossils is explored.

12. Darwin and Dinosaurs. From the study of fossils we have learned how animals and plants have changed with time. Dinosaurs were animals specialized in their adaptations to the conditions in which they lived. A study of the history of dinosaurs provides a good case history for a summary of evolution.

13. Reproduction. One characteristic of living material is that it reproduces, and reproduction is defined as a process by which an animal or plant gives rise to another of its kind. This program is concerned with the general features which are common to all methods of reproduction.

14. Populations and Predators. All living things tend to increase in number. The female codfish, for example, lays over a million eggs and yet the ocean is not filled with cod. Why? They are controlled by competition for food and by predation. This program explores the relation between a population of animals and a predator.

15. Animals Prepare for Winter. Winter is a time of scarcity for animals. Both food and water are hard to come by and in addition the cold presents its own special problems. The various ways in which animals meet this challenge are the subject of this program.

16. Plants in Winter. Plants also have problems during the winter months. Because of the difference in their method of growth and livelihood they are able to use different means to solve the winter problem. This program deals with the adaptations for survival utilized by plants.

17. Now You See It--Now You Don't. The story of animal camouflage is a fascinating one. Animals hide to escape being eaten, but the many ways in which they conceal themselves are as various as the animals themselves.

18. Food From the Sea. The ocean is a vast and unutilized garden. The grass and grazing animals all have their counterparts in the sea. Man is just beginning to develop efficient ways of harvesting the crop from the ocean.

19. Life in a Drop. Borrow a microscope and place a few drops of pond water on a glass slide. Here you will visit a strange new world. Strange forms of many colors can be seen bumping, jerking, and shooting through the water. The microscopic life of concern in this program are the single-celled organisms which are the smallest unit that is alive.

20. Blood--the Internal Ocean. Most animals carry inside them an ocean of water which serves a multitude of functions. In mammals it is circulated through an amazingly complex system of pipes by a relatively simple pump.

21. Plants without Chlorophyll. Some plants are unable to trap the sun's energy because they lack the green pigment, chlorophyll. How these plants survive and where they get their energy are discussed in this program, as well as their use to man.

22. The 40° Miracle--Spring. What effect does temperature have on life? Most animals are dependent on the temperature of the environment: only mammals and birds are free from this restriction. What are the advantages and disadvantages of the warm-blooded and cold-blooded systems?

23. Animal Sounds. Many animals make sounds. Of what use is this to an animal? During this program we see and hear animals that use sound for orientation, communication and social purposes.

24. Amphibians and Reptiles. These are animals that are still very limited to the kinds of environment in which they can live. In this program we see the difference between a reptile and an amphibian and look at some common examples of both.

25. What Flowers are For. Flowers are necessary for the productions of seeds. Some plants are dependent on insects for pollination and they have a great variety of ways to attract insects.

26. Bees. The honeybee shows a tremendous elaboration of instinctive behavior. In this program we see the life of a bee colony. The marvels of the organization of the beehive, the "social" behavior of the bee colony, as well as the ways in which man profits from the bees, are illustrated in this program.

27. Bird Migration. Many birds migrate from the north to the south in the fall and return north with Spring. In this program are discussed the development of bird migration, possible methods by which birds find their way, and some methods used in studying bird migration.

28. The Seashore. This is the first of a group of programs devoted to a survey of various types of communities and the animals and plants that live in each. In each, adaptations of the plants and animals to their environments is stressed.

29. Your Backyard--the Country. In this program we explore some of the different habitats and the animals and plants that live in them. In this country-community of living things, all the principles of the Exploring Nature series are applicable, particularly interdependence of all living things.

30. Your Backyard--the City. This program is devoted to showing that life may be found in the most unlikely places--close at hand. No matter where one lives, he is usually in the company of other living things. Wherever there is food, there will be an animal to eat it. Wherever there is a chance for a plant to grow, there will generally be a plant growing.

Population

The population consisted of 90 fifth grade classes (approximately 2600 children) randomly drawn from 300 fifth grade classes in 30 cooperating communities of urban, suburban and semi-rural composition, within a fifty-mile radius of the city of Boston.

During June, 1959, superintendents of schools in all towns and cities receiving the Channel 2 signal were invited to participate in the study being reported here. Each superintendent was requested, if interested, to supply the following information: (1) population of the community, (2) number of children in fifth grade, (3) number of fifth grades, identified by building, and (4) the number of fifth grades he would be willing to make available for participation in the research project. Using the criteria of per capita expenditure for education 1/ it was determined that the cities and towns were sufficiently similar so that in sampling only the ratio of towns to cities (approximately 3:1) need be considered.

Using a table of random numbers, fifth grades by buildings in the cities and towns were selected, to make a total of 72 experimental classes (48 from the towns, 24 from the cities). An additional 18 classes were randomly selected for controls, using the same sampling procedure, completing the total sample of 90 classes, representing 30 communities. The superintendents were then notified of the results of the above sampling process. Those superintendents who were to supply fifth grade classes for the research were asked to provide that fifth grade class within a given building which had turned up numerically in the sampling process, assuming the fifth grades within a building to be ordered alphabetically by teacher name. They were also asked to forward to the teacher thus involved the following: (1) an explanatory form letter, (2) a flyer describing the Television Workshop, (3) a flyer describing the Science Workshop, (4) a projected schedule of research activities for the month of September, and (5) a return postcard on which the teacher was asked to indicate whether or not she was willing to participate in the project and whether she planned to attend the group meeting for all participating teachers on September 9, 1959 at Boston University.

About 85 per cent of the original sample made good on the superintendents' commitments. During the group meeting for the teachers on September 9, the nature and degree of involvement associated with their participation in the research was explained to them in greater detail. Teachers were assigned to the two workshops on the basis of their expressed preference, filling the quota of 24 for each workshop, but leaving some vacancies to be filled in the "no special training" and "control" groups. These were subsequently filled by repeating the original sampling procedure. At the time of the September 9 meeting, those teachers already committed to the project were given their experimental categories on the basis of random assignment, as well as whatever materials they would need to implement their assignment; e.g., study guides, lists of pupil projects, resumés of program, etc.

1/ Department of Education, Commonwealth of Massachusetts Annual Report for Year Ending June 30, 1958, Part II.

Each teacher was asked to complete a personal data sheet giving information with regard to educational preparation and professional experience in the teaching of science and the nature and extent of the science program in her school; a map or other directions for getting to the schools involved was also requested.

Teacher Assignments

At the time of the initial meeting, participating teachers had the opportunity to meet the staff in charge of the Science and Television Workshops and to plan for their first workshop meeting. The two workshops were handled as extension courses offering graduate credit from Boston University for those who wished it.

NATURAL SCIENCE WORKSHOP FOR ELEMENTARY SCHOOL TEACHERS

This workshop group met for three all-day sessions, September 17, 18 and 19, and then afternoon sessions on Thursdays, September 24 to October 29. The instructional staff was several members of the Massachusetts Audubon Society and all sessions were held at the Nature Center, Drumlin Farm Wildlife Sanctuary, South Lincoln, Massachusetts.

The intensive workshop program covered major biological principles such as the nature of life, reproduction, evolution, animal and plant communities. Emphasis was focused on the elementary school program and projects and materials that can be utilized successfully in the intermediate grades. Field trips were conducted to familiarize teachers with the local resources available.

Topics Covered in the Workshop Sessions

1. Use of Live Animals
2. Energy
3. Reptiles and Amphibians
4. Birds
5. Trees and Shrubs
6. Minerals
7. Marine Life
8. Lower Plants
9. Keeping Animals
10. Three Kingdoms

Topics Covered in the Field Trips

1. Pond Life
2. Bird Migration
3. Insects
4. Vegetation and Succession
5. Geological Deposits
6. Animal Behavior

WORKSHOP IN THE UTILIZATION OF INSTRUCTIONAL TELEVISION PROGRAMS

This workshop group met for three all-day sessions, September 11, 12, and 19, and then afternoons on Wednesdays, September 23 to October 28. The instructional staff was from the Boston University School of Public Relations and Communications and all sessions were held at the SPRC Television Studios.

Designed for those persons interested in utilizing the television medium as a teaching aid in the modern classroom, this course attempted to orient the teacher with the medium and help in the integration of its use into the classroom presentation. Areas covered included the general utilization of audio-visual aids with emphasis on the TV medium, familiarity with equipment and techniques of television, pre- and post- program activities for the classroom, the TV teaching guide, evaluation of instructional television programs in the United States, the needs and future of instructional programs on television, TV in the elementary and secondary curriculum, and other such topics pertinent to successful understanding of the new medium and beneficial use of its offerings in the American classroom today.

Television Workshop Course Outline

1. Introduction: Justification; aims and purposes
2. Background: The problems and complexities of U. S. Education today; the role of modern communications media in education
3. Definition of terms:
    a. open and closed circuit broadcasting
    b. direct and indirect teaching
    c. enrichment, supplementation, replacement
4. TV Production Orientation for understanding and appreciation
5. Studio Demonstrations
6. The Problems of Educational Television are the Problems of Education:
    a. Principles of good teaching
    b. How we learn
    c. The audio-visual aid, its place and value
7. Educational Television History
8. Integration of ETV into Curriculum and Classroom
9. Exemplary types of ETV Projects
10. The Television Teacher
11. Program Viewing and Evaluation; Evaluative Criteria
12. Direct vs. Indirect TV Teaching
13. Effective Use of TV in the Classroom
    a. Class preparation
    b. During the telecast
    c. Follow-up activities
    d. Evaluation
14. Physical Problems of Classroom Use
15. The Teaching Guide and Its Effective Use
16. Research and Results; Future Research Needs
17. Summary and Evaluation

Other Variable Experimental Conditions. Each experimental teacher, plus being in a science-, television-, or no-special-training category, was assigned to a combination of three other variable experimental conditions. These assignments were described to the teachers as follows:

1. Teaching Method

    a. Study guide. Use the study guide (printed and distributed by the 21" Classroom for use in conjunction with the natural science series, EXPLORING NATURE) as a source book in your science instruction with your class. It provides you with essential information on the content of each program which will be helpful in orientation of your class or in follow-up discussions. Questions are provided for use in discussion, plus suggestions for projects. The school or town library will undoubtedly have some of the books suggested as references and it is entirely possible that the children may have some books of their own available.

b. Independent direction. You have been provided with a 21-page resumé of the natural science television programs, which provides a brief overview of each program you and your class will see. Utilize any source material you desire with regard to the programs to be viewed and the topics to be covered. Feel free to use your initiative and imagination in the work you do with your class in conjunction with programs and projects. Do not use or refer to the study guide!

2. Presentation of Programs

a. TV Initiating. Teachers assigned to this category should wait until the first three programs are broadcast before undertaking any science activities with their classes. You will note that the first seven programs focus on animals and their structure, movements, habits, etc. Even the program on plants points up the relationships between plant and animal life. The programs should provide a rich source of stimulation for children's ideas and interests and easily lead into a variety of individual or class projects and assignments. Use these programs as a basis for your work with your class.

b. TV Terminating. In the above category (2a) we see TV as a stimulator of interest. In contrast, for the classes assigned to this category, we see TV as a source of knowledge and a clarifier of ideas. If you have been assigned to a combination involving this category, begin work immediately with your class (about three weeks preceding the first telecast). Look over the content of these first seven programs and encourage your class to find topics related and begin to obtain information from books and other materials available. TV should serve to clarify concepts and provide information on topics which the class has already become interested in and to which they have been oriented.

3. Pupil Activity

a. Individual Projects. Encourage the children, alone or in pairs or trios, to choose topics of interest to them. Their interest may be stimulated by reading a story or seeing a film. Their choice of topics may be promoted and enlarged by preparing a check list of topics which children can check according to their interests. Questionnaires or group discussions can be used to stimulate ideas. A science program being shown on TV after school may be assigned and discussed as preliminary to suggesting individual projects. Those children who complete their regular classwork can be permitted to turn their attention in their free time to their science project as a reward.

b. Class Assignment. We are often reluctant to let children pursue their own interests in a field for fear they won't learn the basic general information. By assigning common topics to all children we know they are more likely to cover the same ground. Whether or not they acquire and retain more basic information than children working alone on individual projects remains to be seen. From the materials available to you, pick topics of probable interest to the class, but assign the same topic to all. They may use different means with respect to the chosen topic, e.g., drawings, talks, models, etc., but all should work on the same topic.

Experimental Design

The following experimental design was effected:

Television Presentation

Teaching Method		TV as Initiating Activity	TV as Terminating Activity	
	STUDY GUIDES	6 classes-Teacher Sci. Train. (3) pupil specialization (3) common class assignment 6 classes-Teacher TV Trained (3) same as above (3) " " " 6 classes-No Spec.Teach. Tr. (3) same as above (3) " " " (18)	6 classes-Teacher Sci. Train. (3) pupil specialization (3) common class assignment 6 classes-Teacher TV Trained (3) same as above (3) " " " 6 classes-No Spec.Teach. Tr. (3) same as above (3) " " " (18)	(36)
	INDEPENDENT DIRECTION	6 classes-Teacher Sci. Train. (3) same (3) " 6 classes-Teacher TV Trained (3) same (3) " 6 classes-No Spec.Teach. Tr. (3) same (3) " (18)	6 classes-Teacher Sci. Train. (3) same (3) " 6 classes-Teacher TV Trained (3) same (3) " 6 classes-No Spec.Teach. Tr. (3) same (3) " (18)	(36)
		(36)	(36)	(72)

Eighteen classes were designated as the control group, receiving the televised natural science series but having no additional class activities. The remaining 72 classes, as can be seen in the foregoing paradigm, were assigned to the following experimental conditions and combinations of conditions:

1. Teacher training. Teachers in the experimental group (72) were assigned to one of the following teacher-training groups, 24 teachers in each group:

    a. Twenty-four teachers participated in a natural science workshop for elementary school teachers conducted by the staff of the Massachusetts Audubon Society. Emphasis was upon the integration of natural science into the elementary school program, methods, materials and projects which can be utilized successfully with elementary school children, and resources available to the teacher. The workshop was the equivalent of a semester extension course and carried graduate credit.

    b. Twenty-four teachers received training in a workshop in the utilization of instructional television programs conducted by the instructional staff of the Boston University School of Public Relations and Communications and was also the equivalent of a semester extension course and carried graduate credit. Emphasis was on the understanding, utilization, integration and evaluation of television as a medium of instruction.

    c. The remaining twenty-four teachers in the experimental group received no special training in conjunction with the study.

2. Teaching method. Each of the teachers in the experimental group was provided with one of two kinds of teaching materials to use with the programs.

    a. One-half of the experimental teachers received a correlated study guide printed and distributed by the 21" Classroom as a source book in conjunction with the natural science series. This study guide provided essential information on the content of each program to be used for class orientation or for program follow-up. Suggested discussion questions and projects were also included.

    b. The other half of the experimental teachers received a brief resumé of each of the programs and were encouraged to use their own initiative and imagination in the work they did with their classes in conjunction with the program series and to utilize as they desired any source material at their disposal. These teachers were specifically instructed not to use or refer to the printed study guide.

3. Presentation of programs. Each of the experimental classes used the television programs in one of two ways, either as an initiating activity or as a terminating activity in relation to classroom study.

    a. In one-half of the experimental classes the weekly television programs preceded related class activities, assignments, projects, etc. Here the programs performed a stimulating or motivating function and served as a basis for subsequent classroom work.

b. In the other half of the experimental classes the weekly programs were preceded by related class activities, assignments, projects, etc. In this instance the function of the programs was to clarify ideas and concepts and to provide information on topics which the class had already become interested in and to which they had been oriented.

4. Pupil activity. In each of the experimental classes, all of the pupils were assigned to one of two types of pupil activities.

a. In one-half of the experimental classes pupil activity took the form of common class assignments given by the teacher to all pupils in the class in conjunction with the natural science programs.

b. In the other half of the experimental classes pupil activity took the form of individual pupil projects undertaken on the basis of personal interests or specialties on the part of the student.

Thus each of the experimental classes was operating under some combination of each of the four experimental conditions, i.e., teacher training, presentation of program, teaching method, and pupil activity.

Testing Schedule

Administration of pretests. Beginning in the middle of September, 1959, and continuing through October, the research staff distributed, administered, and collected the pretest instruments: (1) science information, (2) science vocabulary, (3) standardized science achievement, (4) science reasoning, (5) interest inventory, (6) activity check list, (7) TV viewing check list, (8) attitude scale.

In addition, each teacher was asked to complete the Kuder Vocational Preference Inventory; and, during December, 1959, the Otis Quick-Scoring Beta Test of Mental Ability was administered to all pupil subjects. The occupation of parent, obtained from the permanent school record, for each child was coded according to the following classification system: (1) the Arts--actor, artist, illustrator, author, musician, composer, etc. (2) Scientist--chemist, economist, physicist, psychologist, biologist, astronomer, geologist, etc. (3) the Professions (exclusive of category 2)--accountant, clergyman, teacher, dentist, engineer, lawyer, physician, pharmacist, etc. (4) Other Fields--bookkeeper, clerk or typist, housekeeper, waitress, policeman, firemen, salesman, businessman, farmer, foreman, jeweler, carpenter, machinist, electrician, factory worker, truck driver, baker, etc.

Each teacher was asked to keep a weekly record of the nature of the instruction conducted in the class (vocabulary study, concept development, discussion, readings, audio-visual aids, evaluation of pupil work), numbers and kinds of class assignments or individual projects under way or completed, and the amount of class time devoted to science during the week, exclusive of the time spent watching the program. The teachers in the "control" assignment were requested to have their classes merely watch each of the programs, accompanied by only a brief preview and follow-up discussion, and to engage in no natural science activities or projects for their classes.

Second testing. Some of the instruments, the interest inventory and the tests of achievement, science information and vocabulary, covering the first seven programs were readministered in December, 1959. The data collected at the time, it was hoped, would provide comparative data for short-term evaluation of the effects of the experimental variables on science achievement and interest. The more stable behavior variables such as attitudes and science reasoning ability were presumed not to be susceptible to significant change in so short a space of time; hence these instruments were not employed in the second testing.

Visits to classrooms. During the month of October, schedules for visits to classrooms by the research staff were devised, the objective being to visit each classroom at least once every two weeks, exclusive of the testing periods, and to visit each classroom at least once during the school year when the class was engaged in science activities (except for the control classes) and at least once while the class was viewing the televised natural science programs on Friday mornings.

The program series. The series of 30 natural science programs, entitled EXPLORING NATURE, began with the first telecast on Friday morning October 2, 1959, at 9:15 and repeated at 10:45, telecast times which would prevail throughout the remainder of the school year. A block of the first seven programs was taken as the basis for the first pretest to re-test gain (September to December). The topics dealt with by these seven programs were (1) alive or not alive, (2) these are animals, (3) these are plants, (4) animal tools and weapons, (5) mammals, (6) how animals move, (7) animal homes. An abbreviated science information test was constructed to measure achievement relative to the content of this block of programs and was administered, together with the vocabulary and interest tests, in December, to the 72 experimental classes.

Administration of posttests. The programs were to have been viewed by all 90 classes in the experimental project during all weeks when school was in session (programs were not telecast during public school vacation periods) and the teachers were expected to have adhered to their experimental assignments throughout the school year; however, in February, 1960, during visits to the classrooms the research staff re-emphasized the specific requirements of each of the experimental assignments and encouraged the teachers to make a persistent effort to remain within the limits of the assignment throughout the following weeks in anticipation of the final testing period. During April and May, 1960, each of the instruments administered as a pretest was readministered as a posttest to measure change or gain over the course of the school year for the students in each of the 90 classes in the project.

Treatment of the data. All the instruments used, with the exception of the TV viewing check list, were machine scored. Coding sheets were used to record the test scores for each child, by class, on all tests administered. The data were then transferred to IBM cards, by key punch operation, which in turn were used for analysis of the data by IBM 650 computer, utilizing analysis of variance and correlational techniques. The relationship between each of the experimental variables and the dependent variables was analyzed as well as the interaction between and among experimental variables. The basic unit of comparison was the class not the child. The scores used for analysis, in the main, were mean scores by class with an N of 90 and 18 classes in each cell. The main experimental objective was to determine which classes as a whole showed a significantly greater benefit from the series and under which experimental conditions these gains occurred. In addition, the data collected from the pre- and posttest administrations of the attitude scales, with an original N of 360, were analyzed by individual subjects in an attempt to identify subgroups of children who, irrespective of experimental treatment, showed greatest and least change in the dependent variables and to suggest the particular combinations of factors which might have produced such results.

IV. DATA GATHERED AND RESULTS

Achievement

Test construction. The preparation of test items was preceded by a thorough and systematic examination of the relevant sources for science information and vocabulary. These sources were the TV programs, based on the biological principles as presented by Simpson, Pittendrigh and Tiffany (83); the basic science textbooks with teacher manuals, and appropriate supplementary science readers currently in use within the school systems affiliated with this project.

Cross checking of these various references corroborated the general uniformity of natural science topics covered and provided the content area for both the science information and vocabulary tests. On the basis of the information thus gathered a test blueprint was constructed, with the biological principles listed on one axis and the subject matter of specific programs drafted on the other axis. Test items were then prepared to meet the requirements of this blueprint.

Science Information Test items were designed in the multiple-choice form. Four alternatives were set up from which the answer could be selected. Variations in multiple-choice format were employed. Some sample items follow:

1. Leaves turn toward the
   a. soil
   b. water
   c. warmth
   d. light

Answer Sheet
a b c d
1. ‖ ‖ ‖ ‖

2. Which of the following is fed on milk by the mother?
   a. the hen
   b. the eel
   c. the grey squirrel
   d. the ostrich

Answer Sheet
a b c d
2. " " " "

3. Which of the following does not belong to this group?
   a. the pine tree
   b. the oak tree
   c. the juniper tree
   d. the fir tree

3. " " " "

Six forms of the science information test were constructed from the 190 items developed so that in the preliminary administration of the instruments correspondent coverage of fundamental natural science processes could be achieved and to insure that examinees would have sufficient time to attempt all items given.

Science Vocabulary items were drafted using a word-association technique devised by McLain [1]. Two forms consisting of two sections with seventy-five words to be classified into one of five associative categories, those in the first part describing characteristic groups, those in the second part indicating biological structures and functions.
The following are illustrative of the vocabulary items used:

Section 1

A-Animals
B-Birds
F-Flowers
R-Rocks, ores and minerals
T-Trees

A B F R T
opal
" " " " "
plum
" " " " "
heron
" " " " "
hyacinth
" " " " "
lynx
" " " " "

Section 2

P-Plants
V-Vertebrates
I-Invertebrates
N-Nutrition
R-Reproduction

P V I N R
digestion
" " " " "
fly
" " " " "
hatch
" " " " "
moss
" " " " "
snake
" " " " "

---

[1] McLain, Katherine. The Construction and Evaluation of a Word Association Test of Non-Academic Achievement for Pre-High School Pupils. Unpublished Doctoral Dissertation, Boston University, 1961.

Reading level of instruments. The science information items were carefully developed with regard to reading difficulty and subject matter comprehension. The judgments of elementary school teachers were used as were the opinions expressed in a graduate seminar to determine the suitability of language expression. The vocabulary was checked against the Rinsland and Thorndike 1./ word test to verify subjective evaluation. The readability was found to be well within the comprehension of fifth grade children. To eliminate any possibility of reading difficulty each item on the vocabulary test was read aloud by the examiner during the test administration.

Preliminary administration. To determine item difficulty and validity indices and to detect items which might be weak and ambiguous, a pilot administration of all items was conducted. Selected for the trial administration of the science information test were 650 children in fourth, fifth and sixth grade classes in four representative eastern Massachusetts communities, Forms A to F. The experimental administration of the vocabulary test, Forms A and B, was conducted in fourth through sixth grades using 235 children in a community suburban to Boston. Grades four through six were used to provide sufficient range to encompass the major portion of the ability levels likely to be found in the ultimate fifth grade sample used in the study.

Validity. The validity of the instruments was established in two ways: (1) content validity--the content area to be tested was systematically analyzed to ascertain that all major aspects of the science instruction were covered by the items; and (2) face validity--the judgment of three members of the Massachusetts Audubon Society educational staff, experts in the field of natural science, were obtained as to the appropriateness of the items.

Reliability. Reliability for the six forms of the science information test was determined by the Kuder-Richardson formula for inter-item consistency. Estimated reliability coefficients ranged from .10 to .75 (average .55) for the complete battery. Variance in the reliability coefficients reflected the heterogeneous ability levels of the groups to which the different forms were administered.

Since the final form of the information test contained seventy-five items, the Spearman-Brown Prophecy formula was applied to reliability coefficients obtained on the 30-item preliminary forms to provide an estimate of reliability for the same tests lengthened two and one-half times. Estimated reliability coefficients for 75-item tests ranged from .22 to .88 (averaging .77).

The reliability of the two initial forms of the vocabulary test was determined by applying the Kuder-Richardson formula, estimated reliability coefficients being, for Form A, .91 and, for Form B, .93 (average .92).

1./ Thorndike, E. L. and I. Lorge. A Teacher's Word Book of 30,000 Words. New York: Bureau of Publications, Teachers College Columbia University, 1944.

Item analysis. To determine the difficulty and the discrimination power of items an item analysis was undertaken, using the upper and lower 27 per cent of the cases for the respective tests. The obtained per cents of these extreme groups passing each item were used for computation of item difficulty indices. Phi coefficients were estimated for the proportion of the upper criterion group and the proportion of the lower criterion group passing an item.

Items were grouped according to discrimination power and arranged in order of difficulty. Final selection depended on three qualifications: (1) suitable distribution of item difficulty, (2) highest discrimination and (3) approval by the subject-matter experts.

Results (Achievement). Analyses of variances within and between control and experimental classes at the initiation of the experiment for age, Otis Beta score and I. Q. showed that the groups were comparable in these attributes. Comparisons of experimental and control classes on the initial administration of the science information test (both the complete form and the partial form covering the first seven programs) showed significant differences in favor of the experimental classes. (see Table 1)

Table 1. Analysis of Variance of Science Information Scores Between Control and Experimental Groups Using Total and Partial Means by Class for the First Test Administration.

Data from First Testing	SCIENCE INFORMATION TEST			
	Total Test		Partial Test	
	Mean	S. D.	Mean	S. D.
Experimental Group:	14.39	9.61	7.43	5.67
Control Group:	12.27	9.09	6.40	5.27

Dependent Variable	Source of Variation	df	M. Sq.	F	Confid. Level
Sci. Info (total)	Between Within	1 88	92.720 19.951	4.65	.05
Sci. Info (partial)	Between Within	1 88	26.787 5.532	4.84	.05

Analysis of variance of science vocabulary scores for experimental and control groups revealed no significant differences in vocabulary between these groups at the outset of the experiment. Likewise, analyses of variances of scores for the experimental group, analyzed by experimental treatments, indicated no significant differences among them in the following variables: (1) Age, (2) Otis Beta Score, (3) Otis Beta I. Q., (4) Total Science Information, (5) Partial Science Information, (6) Science Vocabulary. Thus, with only incidental, statistically insignificant differences among them (F values below the .05 level of confidence), the classes assigned to the various experimental conditions within the experimental group could be assumed to be alike in age, intelligence and knowledge of science.

Experimental Teachers. Information obtained from the teachers on length of service and previous training in science and audio-visual education was examined for the purpose of relating these teacher qualifications to the experiment. When years of experience were arranged by the experimental assignment of individual teachers, the range of experience was found to be comparable among the groups.

About eighty-seven per cent of the teachers had nine hours of formal training in elementary school science; the others received some general science instruction in their undergraduate collegiate study. Audio-visual education was quite limited--only about a third of the teachers had at least three semester hours of instruction in this area. For the most part it appeared that the experimental teachers began their experimental assignments with essentially the same backgrounds in these specialized areas.

Second testing. Only the experimental group was involved in the appraisal of the project following the completion of the initial block of seven programs and related classroom activities. It was postulated that this short-term measurement of the experimental effects would not only indicate trends and significances to be further substantiated in the final evaluation, but that it might point to certain distinguishing features between or among sub-groups of subjects. The obtained data were examined for changes in science information and vocabulary among the experimental groups from the first to the second testing.

The difference scores were then treated by a four-way analysis of variances to test the significance of the experimental conditions in relation to gains in the two science areas. (See Table 3). To explore the relationship of sex to differential gains, the same statistical techniques were used with the male and female populations of the experimental classes. Additional analysis was performed to determine whether improvements in achievement were dependent on intelligence and sex, irrespective of experimental assignment. Achievement gains were measured by the partial form of the science information test covering the first seven programs in the series and by the science vocabulary test.

Because the findings from the second testing were, for the most part, repeated and corroborated by the results of the third testing, they will not be reported or discussed separately here. Where warranted, reference will be made to the second testing period data in the summary and discussion which follows the presentation of data for the third testing period, hereinafter referred to as posttest data.

Posttest data. In the final evaluation of this experimental study of the use of television in the teaching of natural science to fifth-graders, experimental and control groups were compared to determine whether television teaching with or without special classroom techniques produced greater gains in the learning of natural science. The effects of the experimental design on differential achievement were analyzed for the purpose of identifying the experimental variables and the interactions of same which appeared most conducive to the effective use of television instruction. Mean differences by class from pre- to posttest measurement of science information and vocabulary were generally the units used in the statistical analyses, with the exception of analyses of the differential gains by intelligence score quartiles where individual scores were employed.

Table 2 shows the results of analyses performed to compare pretest to posttest gains for experimental and control groups in the tests of achievement.

Table 2. Analysis of Variances of Science Information and Vocabulary Achievement Scores from First to Third Testing Between Experimental and Control Groups Using Mean Gains by Class.

Test Period	Dependent Variable	Source of Variation	df	M. Sq.	F	Confid. Level
1 to 3	Sci. Info	Between	1	552.545	19.76	.001
		Within	88			
1 to 3	Vocabulary	Between	1	1271.256	14.35	.001
		Within	88	88.575		

The above relationship of gains to control and experimental classes on the achievement variables was also maintained in analyses performed comparing like sexes within control and experimental groups. Both males and females showed significantly greater gains in the experimental classes than in the control classes in both information and vocabulary measures of achievement.

Analyses of variances performed for experimental classes only, by total class and by sex, using data from both information and vocabulary tests revealed that significant differences in science achievement were not the result of the experimental conditions to which the group was assigned.

Table 3. Four-Way Analysis of Variance of Achievement Gains by Mean Difference Scores for Experimental Classes, First to Second and First to Third Testing Periods.

VARIABLES AND INTERACTIONS	F VALUES											
	SCIENCE INFORMATION						SCIENCE VOCABULARY					
	SEPT - DEC			SEPT - MAY			SEPT - DEC			SEPT - MAY		
	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total
TELEVISION	.10	.07	.05	.01	-	-	.12	.80	.04	2.29	.38	.70
GUIDE	-	.45	.27	.11	.33	.16	.32	.02	.37	1.01	.82	.35
TRAINING	.15	1.20	.52	.43	1.29	.73	.12	.04	.06	.34	.79	.42
ASSIGNMENT	.25	.27	-	.65	.02	.35	-	.30	-	2.00	.01	.17
TVxG	1.56	1.67	1.97	.72	1.57	1.47	.01	.04	.05	-	.20	.51
TVxTr	1.01	.51	.92	.27	.47	.32	.19	.20	.35	.51	.78	.21
TVxA	2.11	8.04*	4.63*	3.63	6.71*	5.07*	2.69	.69	1.82	2.23	-	.45
GxTr	.84	.57	.78	.09	-	.06	.53	.51	.84	.58	.77	.14
GxA	2.48	1.78	2.19	.98	.74	.90	3.36	2.11	3.38	3.63	2.45	2.77
TrxA	.95	.93	.83	.39	1.55	.84	1.66	1.12	1.99	.12	1.60	1.98
TVxGxTr	1.10	.39	.45	.38	.93	.56	.29	.41	.23	1.61	.33	.51
TVxGxA	1.52	3.98*	2.97	.02	-	.01	.69	.03	.36	.54	.14	.05
TVxTrxA	.81	1.12	1.16	1.12	1.38	1.22	2.28	1.67	2.08	4.69*	1.86	2.98
GxTrxA	.17	1.70	1.03	.68	1.73	1.09	.78	.09	.49	.30	1.02	1.96
TVxGxTrxA	3.82	1.22	2.20	2.18	2.09	2.61	1.30	1.91	1.99	.03	2.72	1.80

* significance at .05 level of confidence
** significance at .01 level of confidence

Employing a four-way analysis of variance technique, the four main experimental variables, singly and in all possible combinations with each other, were compared for their contributions to the observed achievement gains from pretest to posttest. (See Table 3). These analyses revealed that in no case could statistically significant gains in achievement, by total class or by sex, be attributed singly to any of the main experimental conditions; i. e., the initiating or terminating use of television, the use of study guides or independent direction by teachers, the specialized training of the teachers, or the use of individual projects or common assignments in classes were not in themselves instrumental in producing significant gains in achievement in natural science as measured by the science information and science vocabulary instruments administered at the onset and at the conclusion of the experiment. In combination, or interaction, however, certain of the experimental conditions did appear to influence gain, chiefly the initiating-terminating use of television together with the individual-common assignment conditions, which produced differential gains significant at the .05 level of confidence among the experimental classes in science information for total group, for males and for females. In vocabulary, differential gains by experimental treatment could be attributed only to the television-training-assignment interaction for females.

One further treatment of the achievement data warrants mention. A random sample of 500 experimental subjects was drawn from the total sample of experimental subjects with complete data. These subjects were ranked by I. Q. scores and these scores divided into quartiles. These I. Q. groups then formed the basis for analyses of variances in achievement by intelligence levels. Table 4 shows the results obtained for analyses of information and vocabulary pretest-posttest data on this basis.

Table 4. Analysis of Variances of Achievement Gains from First to Third Testing Using Means by I. Q. Quartiles for a Random Sample of 500 Subjects Drawn from the Experimental Group.

Test Period	Dependent Variable	Source of Variation	df	M. Sq.	F	Confid. Level
1 to 3	Sci. Info	I. Q. Groups	3	15987.57	123.72	.001
		Trials	1	39262.75	856.14	.001
		Groups x Trials	3	617.08	13.46	.001
		Subjects	496	129.22	2.82	.01
		Residual	496	45.86		
1 to 3	Vocabulary	I. Q. Groups	3	94487.14	117.30	.001
		Trials	1	145467.72	1151.13	.001
		Groups x Trials	3	650.91	5.15	.001
		Subjects	496	805.51	6.37	.001
		Residual	496	126.37		

It can be inferred from the results depicted in Table 4 that the gains observed in achievement by I. Q. quartiles were not attributable to chance factors. The significant interaction value for both information and vocabulary tests implies that gains over trials (pretest to posttest) were not uniform but varied by the level of intelligence. Inspection of the means by I. Q. quartiles reveals that magnitude of gains is related to intelligence level. (See Table 5).

Table 5. Means of Information and Vocabulary Achievement Scores for First and Third Testing Periods by I. Q. Quartiles.

		(LO)	Q_1	Q_2	Q_3	Q_4	(HI)
Science Information	Mean T_1		7.25	12.53	15.00	23.15	
	Mean T_3		15.29	25.22	29.61	37.94	
		(LO)	Q_1	Q_2	Q_3	Q_4	(HI)
Vocabulary	Mean T_1		25.06	37.10	47.47	68.22	
	Mean T_3		44.68	63.65	71.43	94.58	

Classroom instruction and evaluation. Certain methods and techniques typified science instruction in all the experimental classes whether the teacher followed the study guide or developed lessons independently. Examination of the weekly reports filed by the teachers revealed that vocabulary study was the most common method for introducing science topics. Lessons were primarily teacher-centered with general class participation. Charts and bulletin displays were the most frequently used audio-visual aids. Evaluation generally took the form of pupil recitation during question-answer sessions initiated by the teacher. Drawing, reading and collecting were the most common vehicles for the implementation of individual or class science projects or assignments.

Time distribution. Rigid control of instructional time was not feasible because of the scatter of communities involved in the project. The teachers carried out their experimental assignments within the limits of local curriculum provision. From the weekly teacher reports it was determined that the mean times per week devoted to science instruction, exclusive of the telecast time, ranged from 37 to 155 minutes. Pearson product-moment correlations were computed between number of minutes per week devoted to science instruction and the science information and vocabulary pretest-posttest difference scores by class means. The obtained coefficients of correlation were +.21 and +.17 for information and vocabulary gains, respectively. Thus it can be concluded that the relationship between achievement gains and amount of instructional time was positive but slight.

Discussion. Multiple analyses comparing experimental to control classes in achievement gains revealed statistically significant differences in favor of the experimental classes, where the television presentation was accompanied by correlated classroom activities, by total class and by sex. The differences favoring the experimental classes were of greater magnitude in the area of science information than in vocabulary achievement.

Indications discovered at the seven-week interval testing were generally supported by final test data. Differences in achievement gains could not be attributed to the main experimental conditions, teacher training, teaching method, presentation of television, or pupil activity, since each of these variables singly did not contribute significantly to differential gains. The interacting effect of use made of television and type of pupil assignment did, however, produce fairly consistent differences in gains favoring initiating use of television with individual projects and terminating use of television with common class assignment. The relationship of achievement to intelligence level was established using a random sample of subjects with complete data. The subjects with higher intelligence made superior gains to those subjects with lower intelligence regardless of experimental conditions.

The correlation between time devoted to class instruction in science and science achievement was low but positive. Teachers assigned to the various experimental categories were found to be equivalent in experience and training in elementary school science and in audio-visual education.

Table 6. Correlations Between Instruments Used to Study Science Achievement.

		AGE	BETA SCORE	BETA I. Q.	SCI. INFO. Total	SCI. INFO. P_1	SCI. VOCAB. I	SCI. INFO- P_2	SCI. VOCAB. II
CONTROL	Mean	8.70*	33.09	106.63	12.27	6.40	37.78	none	none
	S.D.	7.21	9.40	11.19	9.09	5.27	24.36		
EXPER.	Mean	7.73	35.05	109.18	14.39	7.43	42.93	12.66	55.56
	S.D.	6.29	9.88	10.99	9.61	5.67	24.70	6.60	27.58
AGE	C		-.24	-.57	-.16	-.16	-.15		
	E		-.14	-.46	-.08	-.05	-.08	-.07	-.14
BETA SCORE	C	-.24		.93	.60	.56	.61		
	E	-.14		.93	.63	.55	.65	.56	.66
BETA I. Q.	C	-.57	.93		.57	.54	.57		
	E	-.46	.93		.58	.50	.59	.51	.63
SCI.INFO-Total	C	-.16	.60	.57		.90	.69		
	E	-.08	.63	.58		.91	.73	.66	.68
SCI.INFO - P_1	C	-.16	.56	.54	.90		.61		
	E	-.05	.55	.50	.91		.67	.64	.62
SCI.VOCAB, I	C	-.15	.61	.57	.69	.61			
	E	-.08	.65	.59	.73	.67		.66	.87
SCI.INFO - P_2	C	none							
	E	-.07	.56	.51	.66	.64	.66		.67
SCI.VOCAB, II	C	none							
	E	-.14	.66	.63	.68	.62	.87	.67	

* Age: in months beyond 10.0 years

Inspection of Table 6 reveals moderately high correlation coefficients between science achievement and intelligence and between the different measures of achievement themselves, but low correlations between the achievement measures and age.

Science Reasoning

Test construction. To qualify as a measure of reasoning, an instrument must meet the following criteria:

1. Measure the ability to reason, not test learning or simple recall. For example, contrast the two following items for which the problem is to determine which object does not belong in the group.

   Item 1: (a) carrot, (b) beet, (c) apple, (d) radish

   Item 2: (a) peach, (b) beet, (c) apple, (d) radish

   The first item involves merely a common everyday relationship which the subject may have learned either in school or in day-to-day living. Classifications of fruits and vegetables are so common as to render this item worthless because the answer requires only automatic, simple recall--we know the answer before we start to work the problem. Consider now the second item. The first, automatic classification attempt no longer is successful here. While language habit and memory solved the first item, thought and imagination are required to solve the second. The answer to the second item is that the peach does not belong in the group because the other objects are all red in color; and the validity of this item can be verified by giving subjects a box of colored crayons and asking them to draw each of the objects.

   The important point here is the comparison of the processes involved in the solutions of both problems. Both solutions require memory--the first requires remembering the classification of fruits and vegetables; the second requires the remembering of a commonly abstracted color. The two items differ, however, in that the first is quickly solved by learned habits of language while the second requires thinking.

2. It must not be a test of knowledge or information, per se. This point is illustrated by the following item for which the problem again is to indicate which object does not belong in the group.

Item 3: (a)Kangaroo,(b)Grizzly Bear,(c)Koala Bear,(d)Duck-bill Platypus

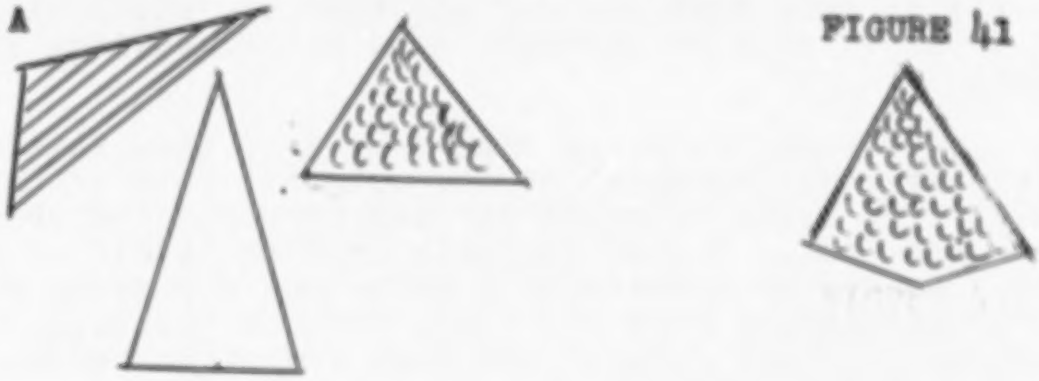
   This item is a reasoning item only if the subject knows that the kangaroo, the koala bear, and the duck-bill platypus are animals native to Australia; and, in that case, the item would favor the individual who had an extensive interest in exotic animals.

3. The third criterion for a reasoning test is that it be able to discriminate between different aspects of the reasoning process, imagination, observation, and language, being the aspects selected for study here.

   a. Imagination. The reasoning process takes place entirely within the nervous system. The information needed to solve the problem is stored in the memory of the subject and the solution of the problem requires the mental manipulation of this information either in the form of verbal symbols, visual symbols, visual images, or any combination of the three. The items contained in the first part of the science reasoning test(referred to as SRI) and illustrated on the previous page, are of this type.

b. Observation. All information required for the solution is inherent in the problem as presented and can be referred to at any time during problem solving. For example, in the item below the problem is to decide whether the figure at the right belongs in the group of figures at the left, having been instructed to observe what the figures in the group have in common. The items in part two of the science reasoning test (SR II) are of this type.

Item: GROUP A FIGURE 41



c. Language. This aspect of the reasoning process consists of using the principles of formal deductive and inductive logic in problem solving. The information necessary to the solution is inherent in the problem and the subject can refer to it at any time during problem solving. This aspect, involving the manipulation of verbal symbols, is similar to the imagination aspect of reasoning except that the solution can be achieved without resort to memory. For example, see the following illustrations of items, the first requiring deductive logic, the second requiring inductive logic. SR III([illegible]rt three of the science reasoning test)contains items of these types.

QUESTION: Do whales nurse their young?

FACTS:
a) Whales are mammals.
b) Baby whales are born alive.
c) Whales are mammals that live in the ocean.
d) Mammals nurse their young.
e) Whales get their air through lungs.

QUESTION: Several people in a particular area became quite sick. Why?

FACTS:
a) Anopheles mosquitos live in a nearby swamp.
b) Grandma Jones says that they have "night sickness."
c) Typhoid makes people sick.
d) Malaria is carried by the anopheles mosquito.
e) Malaria makes people sick.

Preliminary administrations: Science Reasoning Test I. To minimize the effect of previous knowledge on the ability to solve the "imagination" items, each item was drafted first as a recall item and this form of the test was administered to a sample of fourth, fifth and sixth grade children in two Boston suburbs representing a wide range of socio-economic and curriculum-standard levels. After item analysis of the resulting scores, items for the final form were chosen from among those receiving the highest percentage of correct responses, thus insuring the greatest possible common knowledge and making solution of the problem independent of differences in information available to the children.

Another difficulty incurred in the writing of items for this test was the possibility that objects could be classified correctly in more than one way and that odd classifications might occur because of commonly held misconceptions among the children.

To obtain a sufficiently high degree of item validity, in the preliminary administration of the original form of SR I the subjects were asked to write why they grouped the objects in each item as they did. Hence, any item lending itself to multiple classifications or items with popular but erroneous classifications could be eliminated from the final form of the test. In order to appear in the final form of the test each item had to meet the following qualifications:

1. It had to be high on the recall test.
2. It had to be void of multiple or popularly erroneous classifications.
3. It had to have adequate discrimination power. Items were analyzed for difficulty and discrimination, using the top and bottom 27 per cents of fourth, fifth and sixth grade classes from two Boston suburbs.

Science Reasoning Test II. The criteria which had to be met by items included in the test to measure the "observation" aspect of reasoning were:

1. Elimination of the feasibility of well-learned or habitual methods of thinking for problem solution by making these appealing but erroneous distractors.

2. Prior knowledge and experience needed to solve the problem are held to a minimum. The subject, to take this test, needs only to be able to count to six, tell a straight line from a curved line, and to recognize differences in size and shape, all of which are acquisitions well within the experiences of fifth grade children.

A preliminary form of this test was administered to fourth, fifth and sixth grade classes in a Boston suburb and subsequent item analysis using 27 per cent top and bottom criterion points indicated that the majority of the items fell within an acceptable range of difficulty and discrimination power.

Science Reasoning Test III. Preliminary testing of children revealed their ability to arrive at correct solutions to the problems in this test using either deductive or inductive methods and their ability to indicate which facts were used to achieve the solution. While the test is designed around formal and rational methods of using language in thought processes, it is not presumed that the instrument measures the ability of the subject to understand the methods involved. The test indicates which child not only arrives at the "correct" answer but also recognizes the facts which prove his answer.

Two forms of the original test were administered to a fifth and a sixth grade class in a suburban community on a trial basis. The major decision to be made in the use of this test was the method of scoring to be employed. Four possibilities presented themselves: The tests could be scored for (1) the number of correctly written answers to the questions posed (necessarily subjective evaluations), (2) items with all facts correctly marked, (3) the number of facts marked correctly, or (4) the number of facts marked correctly minus the number of facts marked incorrectly. It was assumed that scoring method number one was probably the best measure of the ability to reason; however, the other methods of scoring could be performed by machine and some 3,000 subjects would be the ultimate test population. Correlations between the written test score and the machine method scores based on data from the preliminary administration were computed. It was found that scoring method number four (number of facts right minus number of facts wrong) correlated highly with the hand scoring of the written answers. This fourth method also had other attributes in its favor:

1. It eliminated the possibility of getting a high score by guessing.
2. It gave a greater range of scores than method two, making possible a more reliable measure of reasoning.
3. Although the items were originally chosen for inclusion because they dealt with presumably new problems for the subjects, it became apparent upon administration of the test that some subjects knew the correct answer to the question without using the facts given. In such a case it was important to know if the individual realized which facts supported the answer he gave. This discovery indicates that method four may be an even more valid measure than the written answer.

Item analysis followed the pattern of the scoring methods. Each item was analyzed as a written answer, as a total facts right answer, and as individual facts. Each item thus had two figures for difficulty level and two figures for discrimination ability (one representing the answers, one representing the facts). In addition each fact within each item had a level of difficulty and a discrimination index. Items were selected for inclusion in the final form of the test on the basis of distribution of difficulty levels and discrimination power.

Reliability. To obtain an estimate of the reliability of the three science reasoning tests constructed, a sample of 154 subjects were selected from the experimental population, chosen from seven classes in one of the participating school systems. A restricted sample was preferred to a random sample of all subjects in the experiment in order to provide a rigorous test of reliability. This was the same group used to provide estimates of validity discussed above. The selection process screened the total population of the seven classrooms to eliminate all subjects below the age of 10.0 and above the age of 11.0. Using the Kuder-Richardson formula for estimating reliability, the reliabilities for the three tests were determined using raw score data. (See Table 7).

Table 7. Means, Standard Deviations and Reliabilities of the Three Science Reasoning Tests.

Test		N	M	S.D.	r_{11}
Science Reasoning	I	154	16.0	3.87	.65
" "	II	154	14.35	7.27	.87
" "	III	154	14.85	7.31	.83

Standardized Science Test: In addition to the tests constructed specifically for the project, the Science Test of the Sequential Tests of Educational Progress Battery (STEP test)3/ was employed. This test, according to the manual published by Educational Testing Service, is designed to measure the ability to do scientific thinking rather than to measure the acquisition of science information. The manual lists the types of scientific reasoning believed to be assessed by the STEP test and the percentage of questions designed to measure each type, as follows: 4/

1. Ability to identify and define scientific problems.	10%
2. Ability to suggest or screen hypotheses.	25%
3. Ability to select valid procedures.	17%
4. Ability to interpret data and draw conclusions.	23%
5. Ability to evaluate critically claims or statements made by others.	12%
6. Ability to reason quantitatively and symbolically.	13%

Six fields are represented in the items used in the Science test in the following proportions of the questions: 5/

Biology	40%
Chemistry	16%
Physics	23%
Astronomy	8%
Geology	7%
Meteorology	6%

3/ Sequential Tests of Educational Progress, Cooperative Test Division, Educational Testing Service, Princeton, N. J., 1957.

4/ op. cit., Manual for Interpreting Scores, p. 7

5/ ibid.

(It might be noted at this point that the STEP test contained a higher percentage of "biology" items than other standardized science tests designed for elementary school use--one of the major reasons for its selection for use in this project.)

The STEP test claims content validity only and a reliability coefficient of .91 obtained by Kuder-Richardson formula 20. National norms for fifth and sixth grades based on a combined N of 974 are provided. 6/

Results. Analysis of the pretest data revealed no significant differences between experimental and control classes in performance on Science Reasoning Tests I and III, but did indicate initial significant differences in performance on Science Reasoning Test II and the STEP test. (See Table 8).

Table 8. Analysis of Variances of Pretest Data on Science Reasoning Tests Between Control and Experimental Groups.

Test Period	Dependent Variable	Source of Variation	df	M. Sq.	F	Confid. Level
1st	Sci. Rsn. I	Between	1	10.644	2.64	NS
		Within	88	4.033		
1st	Sci. Rsn. II	Between	1	91.506	4.60	.05
		Within	88	19.905		
1st	Sci. Rsn. III	Between	1	29.987	3.64	NS
		Within	88	8.228		
1st	STEP	Between	1	92.924	4.34	.05
		Within	88	21.414		

Analysis of pretest scores by the classes in the experimental group only revealed no significant differences by experimental assignment in initial performance on Science Reasoning Tests I and III or the STEP test, but a difference significant at the .05 level of confidence was observed in initial performance on Science Reasoning Test II. While a simple analysis of variances in Otis Beta I. Q. scores comparing experimental and control groups by class mean I. Q. showed that differences did not reach a level of significance, a more detailed analysis of differences

---

6/ Technical Report: Sequential Tests of Educational Progress. Cooperative Test Division, Educational Testing Service, 1957, pp. 9, 22, 53.

by sex revealed that what difference was observed was contributed almost entirely by boys in the control group and that the difference between boys in the control and experimental groups measured by Otis Beta I. Q. was significant at the .05 level of confidence. (See Tables 9 and 10 relative to the above discussion.)

Table 9. Analysis of Variances of Pretest Data on Science Reasoning Test II Among Experimental Classes and of I. Q. Data Between Experimental and Control Classes.

Test Period	Group	Dependent Variable	Source of Variation	df	M. Sq.	F	Confid. Level
1st	Exper.	Sci. Rsn. II	Among	23	3351.98	2.05	.05
			Within	48	1633.99		
1st	Exp. vs Control	Otis I. Q.	Between	1	92.0	3.42	NS
			Within	88	26.9		
1st	Exp. vs Control (Boys)	Otis I. Q.	Between	1	169.47	5.01	.05
			Within	88	33.94		

Table 10. Means and Standard Deviations for I. Q. Data and for Science Reasoning Test I Gains, Between Experimental and Control Groups by Total Class and by Sex.

GROUP		N	OTIS Beta I.Q.		SC.RSN.I-Gains	
			Mean	S.D.	Mean	S.D.
TOTAL	Total	90	107.8	5.23	2.38	1.29
	Experimental	72	108.3	5.30	2.53	1.30
	Control	18	105.8	4.39	1.79	1.05
BOYS	Experimental	72	107.9	6.01	2.39	1.86
	Control	18	104.5	4.61	1.26	1.57
GIRLS	Experimental	72	109.7	5.16	2.74	1.46
	Control	18	108.6	4.89	2.47	1.47

Analysis of variances for pretest to posttest gains, using mean difference scores, revealed no significant differences between experimental and control group performance on Science Reasoning Tests II and III or on the STEP Test analyzed by total class and by sex. Analysis of variances for Science Reasoning Test I, however, indicated that between the experimental and control groups there was a difference in mean gain significant at the .05 level of confidence. This significance was contributed almost entirely by differences in gain between boys in the two groups, as can be seen by the data reported in Tables 10 and 11.

Table 11. Analysis of Variances of Gains in Science Reasoning Test I Between Experimental and Control Groups by Total Class and by Sex.

Test Period	Group	Dependent Variable	Source of Variation	df	M.Sq.	F	Confid. Level
1 to 3	Exp vs C (total)	Sci. Rsn. I	Between	1	7.745	4.80	.05
			Within	88	1.612		
1 to 3	Exp vs C (boys)	Sci. Rsn. I	Between	1	18.136	5.43	.05
			Within	88	3.339		
1 to 3	Exp vs C (girls)	Sci. Rsn. I	Between	1	1.035	.47	NS
			Within	88	2.182		

Analysis of variances of pretest to posttest mean differences in gains in science reasoning by experimental assignment, for the experimental group only, produced the following results. In four—way analysis of variances obtained for Science Reasoning Test I significant differences were observed only in mean gains for girls, the four—way combination of TV x Training x Guide x Assignment proving significant at the .05 level of confidence. Similarly, variances obtained for Science Reasoning Test II revealed significant differences in mean gain (.05) for girls only, in the three—way combination of Guide x Training x Assignment. For Science Reasoning Test III, significant differences in mean gain by assignment were obtained for total class, for boys and for girls, consistently in the three—way combination of TV x Training x Assignment. No significant differences in mean gain by experimental assignment were obtained for the STEP test by total class or by sex. (See Table 12)

Table 12. Four-Way Analysis of Variance of Science Reasoning Gains by Mean Difference Scores for Experimental Classes, First to Third Testing Periods.

VARIABLES AND INTERACTIONS	F VALUES FOR SEPT - MAY DIFFERENCES											
	Sci.Reasoning I			Sci.Reasoning II			Sci.Reasoning III			Step Test		
	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total
TELEVISION	.32	3.82	1.47	-	.63	.27	1.15	.04	.38	.54	.05	.33
GUIDE	.65	-	.56	1.69	3.64	3.52	.17	.14	.01	3.18	.87	2.75
TRAINING	.19	1.03	.15	.09	.57	.45	.30	.20	.29	1.13	.14	.97
ASSIGNMENT	2.42	.17	1.26	.03	.06	.08	.11	.20	.04	.63	1.15	.02
TVxG	.72	.35	.02	1.32	.10	.59	.44	-	.35	.45	.04	.53
TVxTr	.45	1.37	1.18	.93	.05	.23	.42	.89	.17	2.61	1.07	2.30
TVxA	1.28	.17	1.78	.03	.41	.14	.83	.78	-	.20	2.54	.36
GxTr	.04	.13	.16	.19	.34	.04	.22	.37	.13	1.24	.87	1.22
GxA	1.69	.30	1.98	1.53	2.98	2.49	3.75	.59	.37	.69	.32	.5[illegible]
TrxA	1.00	1.18	.10	.50	.06	.13	2.77	1.58	2.84	.45	.55	.85
TVxGxTr	1.85	2.43	2.94	1.34	1.65	1.64	.65	.37	.08	.19	.39	-
TVxGxA	.01	1.36	.11	1.16	-	.33	.75	.02	.42	.11	.48	.14
TVxTrxA	.85	.17	.35	.72	.52	.47	[illegible].25*	5.07*	5.01*	.70	.51	.43
GxTrxA	.38	1.03	.89	.84	4.74*	2.63	1.08	1.80	.56	.21	.04	.02
TVxGxTrxA	.32	4.32*	1.43	.01	.38	.16	2.00	.19	.45	.52	1.36	1.51

* significance at .05 level of confidence

Discussion. Small but consistent gains in science reasoning were made by all 90 classes in the study. Only on Science Reasoning Test I, however, was there a significant difference (.05) between experimental and control classes in mean gain, and this significance was contributed by the large difference in mean gain between boys in the two groups. This result becomes interesting in the light of the fact that the difference in Otis Beta I. Q. between experimental and control group boys is significant at the .05 level of confidence.

In general, no significant differences in gain in science reasoning were observed by main experimental effects, and only a few of the experimental treatments in combination influenced gain, and these mainly for girls.

The results raise a question as to what actually brought about gains in science reasoning ability in both the experimental and control classes. While the control classes were not studying biological science in the classroom as were the experimental classes, many of the control classes did have physical science included in their curriculum, and it is conceivable that the control group learned scientific reasoning while studying the physical sciences. Since the experimental group made significantly greater gains in science achievement than the control group, increased ability to do scientific reasoning cannot be attributed to equivalent gains in information in the two groups. Analysis of differences in intelligence could lead to the conclusion that what had been defined as scientific reasoning is closely related to intelligence. However, doubt is cast on this conclusion when the correlations between the reasoning tests and the intelligence test and between the mean gains scores and the intelligence test are examined. (See Table 13). There were indications that individual classrooms were significantly different in their development of scientific reasoning ability, which suggests the possibility that the individual classroom teacher in some way may be affecting the development of scientific reasoning and that boys and girls in the same class may not be affected the same way. There was no evidence as to what might be the source of these differences in the classrooms.

Table 13. Correlations of Instruments Used to Measure Science Reasoning with Otis Beta I. Q., Using Pretest Scores and Difference in Gains Scores.

TEST	CORRELATION WITH I. Q.	
	Pretest Scores	Difference Scores
Sc. Rsn. Test I	.53	.09
Sc. Rsn. Test II	.21	.18
Sc. Rsn. Test III	.55	.24
STEP TEST	.65	N.S.

The STEP test proved of little value in this particular experiment. While it did not yield significant differences between control and experimental groups, between boys and girls, or among experimental treatments, the results obtained from this test did indicate that both the experimental and control classes advanced from pretest to posttest performance to the degree that would be expected on the basis of the published national norms.

Table 14. Correlations Between Instruments Used to Measure Science Reasoning.

		SCIENCE REASONING TESTS			
		I	II	III	STEP
CONTROL	Mean	21.11	13.60	21.12	34.64
	S.D.	5.40	8.73	6.71	10.79
EXPER.	Mean	21.76	15.48	22.08	36.58
	S.D.	5.27	9.63	6.44	10.19
AGE	C	-.20	-.05	-.14	-.21
	E	.11	.01	-.05	.09
OTIS BETA SCORE	C	.56	.35	.56	.66
	E	.56	.23	.60	.71
OTIS I. Q.	C	.55	.32	.52	.63
	E	.53	.21	.55	.65
SCI. RSN. I	C	1.00	.30	.41	.49
	E	1.00	.18	.43	.59
SCI. RSN. II	C	.30	1.00	.42	.29
	E	.18	1.00	.32	.27
SCI. RSN.III	C	.41	.42	1.00	.52
	E	.43	.32	1.00	.55
STEP	C	.49	.29	.52	1.00
	E	.59	.27	.55	1.00

The intercorrelations presented in Table 14 show low correlation coefficients between age and the various science reasoning tests. Moderately high correlations exist between intelligence and Science Reasoning Tests I, III and the STEP test, all with high verbal content, but comparatively low correlations are found between intelligence and Science Reasoning Test II, the only nonverbal reasoning test. The correlations between the reasoning tests themselves range from relatively low to moderately high, the relationship between SR II and the other measures of reasoning being the lowest and those between SR III and the STEP test being the highest.

Interests and Attitudes

Test construction. Failing to find any suitable instruments already in existence, the task of constructing tests to measure interest and attitude status and change was undertaken.
Interest Inventory: A preliminary form of the interest inventory, analogous in design and function to the Kuder Vocational Preference Inventory (Science Research Associates), was drafted. This instrument, containing 44 sets of triple-forced-choice items was designed to inventory expressed interest of subjects in each of five areas: Science, Mechanical, Athletic, Literary and Humanitarian interest. These interest areas were chosen on the assumption that the mechanical and athletic scales would have positive valence for boys, that the humanitarian and literary scales would have positive valence for girls, and that these scales would operate as controls allowing for change in the science score as a function of the effect upon science interest of the experimental conditions obtaining between pretest and posttest administrations of the instrument. The instrument was administered on a trial basis to a fourth, fifth and sixth grade class in June, 1959. With the data thus collected, an item analysis was completed, and as a result of the item analysis, some undiscriminating choices were eliminated entirely; other choices were recombined, and the wording of the items changed in some instances to provide more appropriate fifth grade vocabulary. The items incorporated activities appropriate to the age level of the subjects.

Reliability computed by the Kuder-Richardson formula for determining rational equivalence was found to be .97. Intercorrelations between the five scales, i.e., between total scores rank ordered on items in each of the five interest areas, revealed the following relationships: Science-Mechanical, -.23; Science-Athletic, -.26; Science-Humanitarian, -.49; Science-Literary, -.65. The final form of the interest inventory, entitled "What I Would Rather Do" contained 40 sets of triple-forced choices in which 40 science-interest items were paired alternately with 20 each of Athletic-, Mechanical-, Humanitarian-, and Literary-interest items in the following manner:

(S) 1. a. Visit an aquarium with many kinds of fish
(A) b. Visit a factory where sports equipment is made
(H) c. Go shopping for articles to put in a first-aid kit

(M) 2. d. Design and make model airplanes
(S) e. Raise white mice for scientists
(L) f. Collect rare stamps from all over the world

Activity Check List: Lists of activities in which fifth grade youngsters could be expected to participate were written, incorporating the following main activity areas: Sports, such as baseball, bicycling, swimming; Collecting, such as stamps, butterflies, airplanes, photographs; Other Hobbies, such as making things, exploring; Television or Radio Programs, by type, such as mystery, western, historical; Moving Pictures, also by type; Reading, books, magazines, newspapers; Active Games and Passive Games, by themselves or with others; and Clubs and Organizations, such as Boy Scouts or Girl Scouts, 4-H Club. These lists were so designed that they could be printed on an IBM machine-scorable answer sheet, providing an answer space for each activity. The directions requested the subject to indicate those activities in which he had been engaged at least once during the previous week.

Although the activity check list provides a measure of manifest interest which can be compared with expressed interest as measured by the interest inventory, its primary function was to provide a measure of ascendancy of activities related to television viewing and natural science, and the relationship to these pursuits of the experimental treatments.

TV Check List: A check list containing a representative sample of late-afternoon and evening program telecast during any two days of randomly selected weeks during the school year was constructed and administered once during the pretest period in the fall and once during the posttesting in the spring. The response made to each program listed was in terms of whether or not the subject had watched that program and whether he watched because (1) his parents suggested that he watch it, (2) he watched it by chance, or (3) he planned to watch it. The TV check list was designed to provide data on the amount and kind of viewing done by the subjects; the relationship of the subject's interests and leisure-time activities to his television habits. It was thought that televiewing habits (kind of program, frequency and duration of viewing) could have implications for the effect upon the subject's behavior of the program series in natural science.

Attitude Scale: In June,1959, the following sentence-completion test was administered to a class of fourth, fifth and sixth graders, respectively:

READ THE FOLLOWING SENTENCES AND WRITE DOWN WHAT YOU THINK OF RIGHT NOW:

1. When I grow up I would like work in which I ______________________
2. When I think of a businessman I think of ______________________
3. When I think of a scientist I think of ______________________
4. When I think of a teacher I think of ______________________
5. When I think of a mechanic I think of ______________________
6. The thing I like most about scientists is ______________________
7. The thing I dislike most about scientists is ______________________

The responses given to these incomplete sentences (numbers 3, 6, and 7) were found to fall closely within the categories (positive, negative and shared images) of perceptions of scientists found by Mead and Metraux (67:387-9) to be prevalent among high school students (N= 35,000). The scientist was seen by the adolescents as a near-genius who studies and works very hard, sometimes for days without food or sleep, carefully conducting experiments, often with repeated failure and frustration, motivated not by desire for money or fame but for the benefit of mankind. He is perceived as a "brain" engaged in dangerous or secret work, neglecting his wife and children as well as himself, lacking friends, hobbies or recreation other than a book. Scientists are seen as essential people for society, even if they are elderly, stooped, surrounded by animals (mostly dead) and equipment and perennially pouring chemicals or peering through microscopes or telescopes; but they are nothing to become or to marry, for it means sacrifice of early marriage, early parenthood, early enjoyment of adult life, with career precluding the prevalent values of American life--human relations, family, group membership and participation.

A preliminary form of an attitude scale was constructed containing 66 statements, each prefixed by the phrase, "A SCIENTIST IS SOMEONE WHO.......," to which the subject was to respond first with TRUE, FALSE, OR DON'T KNOW, and then with GOOD, BAD, OR DOESN'T MATTER. The perceptions of scientists by adolescents as reported by Mead and Metraux and the responses of the elementary school children to the sentence-completion test were the basis for the construction of the original 66 statements. Approximately one-half of the statements were designed to correspond to positive (favorable) images of scientists and one-half to correspond to negative (unfavorable) images of scientists, but presumably any or all of the statements could elicit a "Don't Know" or a "Doesn't Matter" response as well as a positive or negative response.

After administration of the original scale to a small sample of adults and children, it was deemed advisable to present the attitude scale in two forms: Part I-A, to measure values or attitudes, to which the child responded by indicating whether he thought the activity represented to be good, bad, or of no matter; and Part II-B, to measure perceptions or beliefs about scientists, to which the child responded by indicating whether he thought the activity or description to be true or false of scientists, or that he did not know.

In addition some statements of dubious significance were eliminated and other statements were edited, resulting in a final 60-item form of the attitude scale in two parts: Part I-A containing statements about people and things that people sometimes do, written in the active form,

		G	B	DM
1.	working mostly with living things	"	"	"
2.	experimenting with animals	"	"	"
3.	usually succeeding in what one does	"	"	"
4.	having just about the most important job in the country	"	"	"
5.	having some pretty odd ideas at times;	"	"	"
		-	-	-

and Part II-B containing statements carrying the same connotations as in the first part, prefaced by the phrase, "A SCIENTIST IN SOMEONE WHO.......,"

A SCIENTIST IS SOMEONE WHO.....		T	F	DK
1.	works mostly with living things	"	"	"
2.	experiments with animals	"	"	"
3.	usually succeeds in what he does	"	"	"
4.	has just about the most important job in the country	"	"	"
5.	has some pretty odd ideas at times.	"	"	"
		-	-	-

Each part of the scale (60 items with the appropriate answer space designation) was printed on a separate IBM machine scorable answer sheet.

Determination of "Objective Reality." To obtain a reference point with regard to what constructs could legitimately be used to describe scientists and their work, against which children's perceptions could be measured, a random sample of 300 scientists (100 each from the physical, biological and behavioral sciences) were drawn from the 9th Edition of American Men of Science [7] and asked to respond to Part II-B of the Attitude Scale. Those responses, indicating what was true or false of scientists as a group, on which there was two-thirds agreement among all the scientists in the sample (about 70% of the items) were used to determine the degree to which the children's responses approximated or deviated from the "reality" provided by the scientists themselves.

Determination of Cultural Values and Stereotype of Scientists. A random sample of 200 adults stratified by occupational levels as classified by the Dictionary of Occupational Titles [8] were administered the Attitude Scale, Parts I-A and II-B, to determine reference points for prevailing cultural values (what is good, bad or doesn't matter) and cultural stereotype of scientists (what is true, false or indeterminate). Those responses on which there was two-thirds agreement among the six occupational levels ranging from professional to unskilled, corresponding in size of sample to the proportion of each occupation in the general population, were used to determine the degree to which the children's responses approximated or deviated from the prevailing cultural value judgments and cultural stereotype of scientists in the adult population.

Function of the attitude scales and rationale for their construction. The attitude scales were designed for the following functions:

1. To determine the degree to which the attitudes of the children towards science and scientists were positive, negative or neutral.
2. To determine the degree to which the children's perceptions of scientists corresponded to reality.
3. To determine the degree of shift occurring in attitude and perception during the course of the experimental period and its relationship to
    a. The educational experiences being provided

---

[7] Cattell, Jaques (ed.), American Men of Science: A Biographical Directory. New York: R. R. Bowker Co., 1955-6. (Ninth Edition, 3 volumes).

[8] Dictionary of Occupational Titles, Division of Occupational Analysis, U. S. Employment Service. Washington: U. S. Government Printing Office, 1949.

b. The similarity of children's perceptions or beliefs to the stereotypes characterizing adults
c. The similarity of children's values to the values characterizing adults
d. The persistent (unchanging) attitudes held by children themselves.

The assumption is made that change in attitude will be a function not only of the experimental treatment of the subject but also a dynamic function of the subject's original attitude structure (identified scale type):

1. attitude type:

   a. positive--ready to perceive, receive, react, accept, approve positively loaded statements

   b. negative--reluctant to perceive, receive, react, accept, approve positively loaded statements

   c. uncertain--nonselective or randomly selective of stimuli; inability or slowness to react due to lack of ready response and intellectual difficulty or lack of understanding; however, where the information is meaningful and comprehension is probable, there should be an increase in belief in the direction of the informational emphasis or loading (positive or negative, favorable or unfavorable), and in some instances the manifestation of a developing attitude, again in the direction of the emphasis or loading of the presentation.

2. cumulative lack of agreement (discrepancy) between responses representing the following comparisons:

   a. objective reality with children's perceptions of science and scientists

   b. children's perceptions or beliefs with their value judgments

   c. children's perceptions or beliefs with cultural (adult) stereotype of scientists

   d. children's value judgments with cultural (adult) values

3. hypothetical predictive scale type:

   a. flexible, fluid--positively correlated with high discrepancy between child's score and the various reference points used for comparison.

   b. vacillating--no correlation with discrepancy between child's score and the various reference points used for comparison.

c. stable, rigid--negatively correlated with high degree of discrepancy between child's score and the various references points used for comparison, e.g. scientists, adults, peers. Stability is conceptualized as a function of low discrepancy between child's perceptions or beliefs and his values and between child's perceptions and objective reality. Rigidity is conceptualized as a function of low discrepancy between child's perception and adult stereotype regarding science and scientists and between child's value judgments and adult values. Both stability and rigidity are hypothesized to be developmental functions of experience and learning, therefore correlated with age and education as well as with certain idiographic data such as religious and ethnic affiliation, area of residence, parental occupation, and probably correlated with intelligence and certain personality factors; the investigation of these hypotheses, however, (with the exception of parental occupation) does not lie within the scope of this particular study.

Results. Analysis of the pretest data indicated no significant differences between experimental and control groups in leisure activities, in responses to Attitude Scale II-B scored with a key for scientists' perceptions of themselves, or in responses to Attitude Scale II-B scored for positive image of scientists. In responses to the Interest Inventory and to Attitude Scale II-B scored with a key for negative images, statistically significant differences at the .05 level of confidence were obtained between experimental and control groups, indicating that the experimental group had a somewhat higher initial interest in science and a somewhat less negative image of science and scientists than did the control group.

The interest inventory was readministered to the experimental classes only at the second testing period. The short-range results revealed no significant changes in interest among experimental conditions. Analysis of variance of interest using a random sample of 266 boys and 234 girls from the experimental group showed a highly significant difference in interests (.01) existed by sex, but that interest changes by sex from pretesting to second testing were not significantly different for boys and girls. Using this same random sample of 500 experimental subjects, analysis of variance of interest from pretesting to second testing by I. Q. quartiles revealed that there were significant differences in interest changes among the intelligence levels, the higher I. Q. quarters tending to decrease in science interest and the lower I. Q. quarters tending to increase in science interest as measured by the interest inventory.

Posttest data. Differences between experimental and control classes in pretest to posttest change in leisure activities and in responses to Attitude Scale II-B scored with the key for scientists' perception of themselves did not reach the level of statistical significance when analyzed by total class or by sex. Changes in interest in science were significantly different for experimental and control girls, the experimental subjects showing a loss from initial interest and the control subjects showing an increase over initial interest. By total class or by boys, analysis of variances between experimental and control groups did not prove significantly different.

Analysis of variance of science interest by I. Q. quarter for pretest to posttest change corroborated the indication of the second testing period results. There was a statistically significant difference in interest changes by I. Q. quarter, the greatest loss of interest in science occurring in the two highest I. Q. quarters while the least loss of interest occurred in the lowest I. Q. quarter, and no change in interest occurred in the next to the lowest quarter. The over-all picture is one of loss of interest in science with the greater losses being associated with the higher intelligence groups. (See Tables 15 and 16)

Analysis of variance of leisure activities by intelligence revealed significant differences in activities chosen from the pretest to the posttest (.01) and among subjects (.01), but differences were not significant by I. Q. levels. (See Tables 15 and 16). Similar analysis by intelligence of variences of Attitude Scale II-B pretest to posttest, indicated a consistent loss in realistic perception of scientists over all I. Q. levels with the largest losses occurring in the two highest intelligence groups. Although differences between I. Q. groups was significant at the .05 level of confidence and the difference between pretest and posttest performance was significant at the .01 level of confidence, analysis of variances by groups over trials fell far short of significance. (See Tables 15 and 16)

Table 15. Mean Scores by I. Q. Quartiles for First and Third Testing Periods on Instruments to Measure Interests, Activities and Perception of Scientists.

I. Q. Quartiles (HI)	INTEREST		ACTIVITIES		II-B PERCEPT.	
	Pre	Post	Pre	Post	Pre	Post
Q-1	14.40	13.32	9.29	8.69	16.44	15.30
Q-2	12.71	11.55	8.65	7.78	15.10	13.63
Q-3	12.53	12.53	9.61	8.90	14.94	14.16
Q-4	12.16	12.09	9.87	8.47	14.85	13.98
(LO)						

Table 16. F-Values for Analysis of Variances by I. Q. Quartiles Over Testing Periods for Interest, Activities, and Perception of Scientists.

Source of Variation	df	INTEREST	ACTIVITIES	PERCEPTION
I.Q. Groups	3	3.48*	1.10	3.28*
Trials pre-post	1	6.41*	9.56**	19.17**
G x T	3	1.89	.37	.41
Subjects	496	3.72**	2.45**	2.66**

* .05 Confidence level
** .01 Confidence level

Four-way analysis of variance by experimental assignment using mean difference scores, pretest to posttest, revealed the following differences were significant for the experimental group: (See Table 17)

1. Only one main effect, teacher training, proved effective in providing significant differences among experimental treatments. This occurred for girls on the Activity Check List, significant at the .05 level of confidence.
2. Three two-way combinations of experimental treatments provided significant differences among experimental treatments. TV presentation by teacher training produced differences (.05) on the Interest Inventory for total class and for boys, but not for girls. TV presentation by pupil assignment provided a significant difference (.01) for girls on the Interest Inventory; also study guide by assignment was significant for girls on the Interest Inventory (.05).
3. The only three-way combination of treatment to produce a significant difference in performance was TV presentation by study guide by teacher training, for boys only, on the Attitude Scale II-B, perception of scientists (.05).

The relationship of science interest to science-related television viewing was studied by (1) drawing the 100 highest and 100 lowest interest scores from pretest data on science interest and comparing these with scores obtained by the same subjects on the TV viewing check list, (2) by comparing the original interest scores of these same subjects with the posttest scores obtained by the same subjects on the TV viewing check list, and (3) by drawing the 100 highest and 100 lowest interest scores from posttest data on science interest and comparing these with posttest scores obtained by the same subjects on the TV viewing check list. Computation of significance was by means of phi coefficients converted to tetrachoric correlations. As might be anticipated, the children's television viewing outside of class was related to their interest in science, those with high measured interest viewing more science related television programs and those with low interest in science viewing significantly fewer science-related television programs. The data would support the conclusion that this tendency remained consistent throughout the year. (See Table 18)

Table 17. Four-Way Analysis of Variance of Changes in Interest, Activities, and Attitude (Part II-B: Perception of Scientists) by Mean Difference Scores for Experimental Classes First to Second and First to Third Teaching Periods

VARIABLES AND INTERACTIONS	F VALUES											
	INTEREST						ACTIVITIES			ATTITUDE		
	Sept - Dec			Sept - May			Sept - May			Sept - May		
	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total
TELEVISION	.07	.32	.47	.11	.64	.35	1.38	.22	.77	.05	1.53	.37
GUIDE	.19	.55	.10	1.42	1.78	.14	.57	.02	.06	.03	1.85	.57
TRAINING	.31	.79	.07	1.64	.40	.38	2.39	3.43*	3.15	2.16	.52	.47
ASSIGNMENT	1.06	1.46	1.26	.10	.77	.05	.52	.03	.14	1.07	.33	.80
TVxG	.09	.87	.05	-	.24	.04	.09	-	-	.44	.10	.01
TVxTr	2.40	1.26	2.68	4.15*	1.61	3.94*	.01	1.31	.42	.24	.23	.04
TVxA	1.79	.01	.21	.06	7.82**	2.70	-	.01	-	.18	1.15	.34
GxTr	1.92	.94	.93	.38	.54	.06	.31	.23	.06	1.24	.89	1.07
GxA	.97	2.80	1.76	.56	4.84*	3.10	.33	-	.09	.18	1.18	.11
TrxA	.79	1.18	1.06	.25	.73	.64	.61	3.66	1.37	.51	3.54	2.13
TVxGxTr	.15	.21	.14	1.04	.09	.62	.24	.76	.29	4.37*	.48	1.82
TVxGxA	.69	.00	.56	.01	1.35	.33	-	.66	.11	.27	.08	.06
TVxTrxA	2.09	.27	.63	3.55	.32	2.75	.72	1.62	1.09	.63	.11	.31
GxTrxA	.05	1.93	.64	.09	.16	.37	.06	.24	.07	.06	.15	.02
TVxGxTrxA	.57	1.97	1.32	.23	.11	.45	.45	.20	.33	.18	2.54	.89

* significance at the .05 level of confidence
** significance at the .01 level of confidence

Table 18. Relationship of High and Low Science Interest to Science-Related Television Viewing Employing Phi Coefficients Converted to Tetrachoric Correlations.

Test Period	Groups	Variables	df	ϕ	r_t	Confid. Level
1-1	HI vs LO	Interest-TV viewing	1	.18	.28	.01
1-3	HI vs LO	Interest-TV viewing	1	.16	.25	.01
3-3	HI vs LO	Interest-TV viewing	1	.32	.48	.01

A random sample (N=360) of Parts I-A and II-B of the Attitude Scale were drawn from the 90 classes in the study for analysis by individuals. The tests thus drawn were scored and divided into top and bottom quarters on the basis of subject's certainty-uncertainty of attitude and perception of scientists and on the basis of correspondence or lack of correspondence to values and perceptions expressed by adolescents, by adults, by scientists, and by their own peers. Status on these variables in turn became the basis for analyzing changes in attitude toward science and scientists and in realism of perception of scientists. Analysis of variance and multiple regression analysis were the methods employed with this portion of the data, using the individual child (with high or low status on the predictor variables) as the unit of study. Table 20 shows the results of the analyses of variances.

Using 269 subjects, those with complete data from the original sample of 360 (4 subjects randomly drawn from each of 90 classes), multiple regression analyses were computed for the two critical variables, attitude toward and perception of science and scientists, yielding multiple coefficients of correlation of .56 and .40, respectively. The significance of these values, with 269 degrees of freedom less a total of eight variables, is in excess of the .01 fiducial limit for the maximum value of R which might occur due to chance. Since the obtained R values of .56 and .40 exceeded the criterion value for significance, the multiple association of the six predictor variables may be considered significant for prediction of change in attitude and in reality perception as herein defined. (See Table 19)

Table 19. Partial and Multiple Correlations of Critical Variables of Perception of Scientists and Attitude Toward Science and Scientists with Six Predictor Variables Related to Children's Values and Beliefs.

Predictor Variables	N=269 Mean	S.D.	Critical Variables PERCEPTION Partial r_{1j}	ATTITUDE Partial r_{1j}
Adult Stereotype	20.30	5.59	.10	.17
Adult Values	31.12	5.39	.05	.05
Adoles. Positive Image	20.78	6.45	.03	.06
Adoles. Negative Image	18.10	6.69	.06	.08
Uncertainty	20.98	10.53	.05	.01
Stable Peer Values	30.33	4.87	.05	.17
	Multiple Correlation:		.40	.56
	Confidence Level:		.01	.01

Table 20. Analyses of Variances of Changes in Perception of Scientists and of Consistency of Beliefs and Values (Attitude) by Status of Subjects on Predictor Variables Derived from Attitude Scale Parts I-A and II-B.

Predictor Variable	Group	Dependent Variable	Source of Variation	df	M.Sq.	F	Confid. Level
Adult Stereotype	Quarters: HI vs LO	Reality Perception	Between Within	1 132	82.26 2.44	33.71	.01
Adult Values	HI vs LO	Reality Perception	Between Within	1 132	.55 2.79	.20	NS
Positive Image	HI vs LO	Reality Perception	Between Within	1 132	102.71 2.98	34.47	.01
Negative Image	HI vs LO	Reality Perception	Between Within	1 132	39.83 3.06	13.02	.01
Uncertainty Re:Scientists	HI vs LO	Reality Perception	Between Within	1 131	76.95 2.30	33.46	.01
Peer Values	HI vs LO	Reality Perception	Between Within	1 132	1.02 2.72	.38	NS
Adult Stereotype	HI vs LO	Attitude	Between Within	1 132	156.55 6.42	24.38	.01
Adult Values	HI vs LO	Attitude	Between Within	1 132	34.81 7.54	4.62	.05
Positive Image	HI vs LO	Attitude	Between Within	1 132	241.61 6.77	35.69	.01
Negative Image	HI vs LO	Attitude	Between Within	1 132	52.01 6.30	8.26	.01
Uncertainty Re:Scientists	HI vs LO	Attitude	Between Within	1 131	31.68 6.79	4.67	.05
Peer Values	HI vs LO	Attitude	Between Within	1 132	49.68 6.67	7.45	.01

It can be observed from the above data (Table 20) that a subject's position in the top or bottom quarter of the group on measures of the predictor variables generally made a statistically significant difference in the changes obtaining from pretest to posttest in the dependent variables. As a further check on the validity of these relationships, multiple regression coefficients were computed for both of the critical variables with all of the predictor variables. (See Table 19)

Table 21. Correlations Between Instruments Used to Measure Interests and Attitudes.

		AGE	BETA SCORE	I. Q.	INTEREST (1st)	INTEREST (2nd)	ADOL. POS.	ADOL. NEG.	II-B SCI. PER.	ACTIVITIES
CONTROL	Mean	8.70	33.09	106.63	12.18	none	21.16	19.74	15.29	8.71
	S.D.	7.21	9.40	11.19	5.26		6.00	6.23	4.59	6.52
EXPER.	Mean	7.73	35.05	109.18	12.83	13.27	21.08	18.71	15.41	9.51
	S.D.	6.29	9.88	10.99	5.30	10.19	6.16	6.41	4.79	6.30
AGE	C	1.00	-.24	-.57	.03		.06	.21	.01	.06
	E	1.00	-.14	-.46	-.06	-.01	.04	.05	-.00	-.04
BETA SCORE	C	-.24	1.00	.93	.06		-.10	-.26	.03	-.04
	E	-.14	1.00	.93	.17	-.00	-.02	-.13	.11	-.02
I. Q.	C	-.57	.93	1.00	.04		-.10	-.29	.02	-.05
	E	-.46	.93	1.00	.17	-.00	-.03	-.13	.10	.02
INTEREST(1st)	C	.03	.06	.04	1.00		.12	.10	.11	.13
	E	-.06	.17	.17	1.00	.35	.08	.07	.07	.25
INTEREST(2nd)	C									
	E	-.01	-.00	-.00	.35	1.00	.07	.05	.01	.20
ADOL. POS.	C	.06	-.10	-.10	.12		1.00	.34	.88	.19
	E	.04	-.02	-.03	.08	.07	1.00	.20	.86	.12
ADOL. NEG.	C	.21	-.26	-.29	.11		.34	1.00	.27	.07
	E	.05	-.13	-.13	.07	.05	.20	1.00	.10	.06
II-B SCI.Per.	C	.01	.03	.02	.11		.88	.27	1.00	.13
	E	-.00	.11	.10	.07	.01	.86	.10	1.00	.10
ACTIVITIES	C	.06	-.04	-.05	.13		.19	.07	.13	1.00
	E	-.05	-.02	.02	.25	.20	.12	.06	.10	1.00

Table 21 shows that little relationship was found between measures of age and intelligence and the various measures of interest and attitude employed. Correlations between interest and attitude as measured were relatively low. Noteworthy is the high correlation between children's perception of scientists and the positive image of scientists held by adolescents.

Table 22. Correlations Between Predictor and Criterion Variables Used to Measure Changes in Attitude Toward and Perception of Science and Scientists.

	I. Q.	$II-B_1$ SCI. PER.	$II-B_2$ SCI. PER.	IA_1 - IIB_1	IA_2 - IIB_2	UNCERTAINTY	STABLE VALUES	ADULT II-B	ADULT I-A	ADOL. POS.	ADOL. NEG.
Mean	109.15	15.23	14.34	9.27	4.88	20.99	30.34	20.78	18.10	20.31	31.13
S.D.	11.92	4.71	5.57	8.24	6.12	10.53	4.87	6.45	6.69	5.60	5.39
I. Q.	1.00	.12	-.06	.08	-.10	.01	.20	.02	-.04	.10	.23
$II-B_1$ SCI. PER.	.12	1.00	.44	.52	.10	-.69	.04	.91	.22	.90	.08
$II-B_2$ SCI. PER.	-.06	.44	1.00	.30	.21	-.24	-.00	.40	.00	.39	.04
IA_1 - IIB_1	.08	.52	.30	1.00	.34	-.28	.23	.56	-.10	.41	.15
IA_2 - IIB_2	-.10	.10	.21	.34	1.00	-.21	.05	.21	.13	.12	.00
Uncertainty	.01	-.69	-.24	-.28	-.21	1.00	-.08	-.78	-.81	-.82	-.11
STABLE VALUES	.20	.04	-.00	.23	.05	-.08	1.00	.03	.09	.05	.83
ADULT II-B	.02	.91	.40	.56	.21	-.78	.03	1.00	.27		
ADULT I-A	-.04	.22	.00	-.10	.13	-.81	.09	.27	1.00		
ADOL. POS.	.10	.90	.39	.41	.12	-.82	.05			1.00	.13
ADOL. NEG.	.23	.08	.04	.15	.00	-.11	.83			.13	1.00

Intercorrelations between Otis I. Q. and the variables used to measure attitude change, as shown in Table 22, indicate little relationship. Correlations between the measures of perception and attitude range from lows approaching zero to highs of .90.

Notably high relationships were found between children's perception of scientists and adult stereotype of scientists, and between children's perception and positive or favorable adolescent image of the scientist. Stable values (unchanging) of the children were also highly correlated with negative or unfavorable adolescent image of the scientist. Degree of uncertainty expressed by the children with respect to beliefs about scientists was highly correlated (negatively) with adult values and stereotype as well as with positive adolescent image. It is also worthy of note that a moderately high correlation between consistency of children's values and beliefs (positive or negative attitude), represented in Table 22 by IA - IIB first and second administrations, and perception of scientists and between consistency of attitude and adult stereotype computed from the pretest data are found to become low correlations, when computed from posttest data.

Discussion. Analysis of the pretest data showed the classes used in the study to be generally comparable. The experimental group had a slightly higher interest in science at the outset than did the control group. There was no difference between the groups in positive image of scientists (the control group had a slightly more negative image than the experimental group); no difference in their realism of perception of scientists, nor in their choices of leisure activities.

Interest in science, measured at mid-year, was developing a trend which remained consistent at the end-of-year testing. The boys had higher interest in science than the girls initially, and their interest increased slightly by mid-year while the girls showed a loss in interest. Analyzed by intelligence, science interest was initially highest in the top I. Q. quarter and lowest in the bottom I. Q. quarter. By mid-year the bottom quarter is shown to have increased interest but the top quarter shows a loss in interest. No differences in interest were found to exist between control and experimental classes as a whole; however, comparing control and experimental girls only, the experimental subjects showed a slight loss from the beginning to the end of the year and the control subjects showed a slight gain in science interest. When treated by experimental assignment, the data reveals there is a significant interaction between teacher training and use of television for initiating or terminating study affecting interest change. The greater losses in interest occurred where the teachers had participated in the science workshop, used individual assignments, had the study guide available and used the programs to initiate the study of science, a combination which suggests that loss of interest may be related to the degree of emphasis put on science by the teacher and the amount of work demanded of the children in connection with the project.

There was no significant difference in leisure activities between control and experimental classes, nor among experimental categories for total classes. However, girls in the experimental classes led by science-trained teachers utilizing individual assignments showed a decrease in science-related leisure activities whereas girls working on individual assignments in classes led by teachers without special training showed an increase. Analysis by intelligence shows a decrease in science-related activities at all intelligence levels.

The relationship between science interest and science-related television viewing was shown to be significant for high and low interest subjects both at the beginning of the experimental period and at the end. The children with high science interest consistently watch more science-related television programs.

There was no difference in the realistic perception of scientists between the control and experimental groups from first to third testing, by total class or by sex; however, by experimental category by sex the interaction effect of television, study guide and teacher training was significant for boys and of teacher training and assignment for girls. Analyzed by I. Q., differences in reality perception from first to third test period are found to be significant among I. Q. quarters, between trials, and among subjects; in each instance the mean values by I. Q. quarter decreases from first to third testing period, indicating a loss in realism of perception of scientists.

Change in the realistic perception of scientists analyzed as a function of the similarity of subject's beliefs about scientists to the adult stereotype of scientists is highly significant, with the high quarter in similarity to adult stereotype showing a decrease in realism and the low quarter a gain in realism from first to third testing period.

The value obtained from comparison of children's responses to Parts I-A and II-B for both first and third testing period is treated as a measure of consistency (where "good" is also "true" and "bad" is also "false") as well as a measure of attitude (positive or negative values). Change in consistency of beliefs and values and in attitude as a function of similarity of subject's beliefs to the prevalent adult stereotype of scientists is shown by analysis to be highly significant; those subjects with high similarity of beliefs to the adult stereotype of scientists show a marked decrease in consistency of beliefs with values and increased negative attitude toward scientists. Those low in similarity to adult stereotype show little change.

Change in reality perception of scientists analyzed as a function of the similarity of the subject's values to the adult value system yields no significant differences between those high and low in similarity to adult values.

Analysis of change in consistency of beliefs and values and in attitude toward scientists as a function of the similarity of the subject's values to the prevalent adult value system shows a significant difference between those in the high and low quarters on similarity. There is a tendency for both groups to move in the negative direction, but the loss is significantly greater for the subjects with high similarity to adult values.

Analysis of change in reality perception as a function of similarity to the commonly held adolescent positive image of the scientist proves highly significant. Those subjects in the top quarter on similarity exhibit a marked decrease in reality perception while those in the bottom quarter on similarity show a significant increase.

Analysis of change in consistency of beliefs and values and in attitude toward scientists as a function of the similarity of the subject's beliefs to the commonly held adolescent positive image of the scientist proves highly significant insofar as the subjects in the low quarter on similarity show little change but the subjects in the high quarter on similarity to the adolescent positive image exhibit a large decrease in consistency and considerable movement in a negative direction in attitude toward scientists.

Change in reality perception analyzed as a function of the similarity of the subject's beliefs to the commonly held adolescent negative image of the scientist proves to be a highly significant relationship. Those subjects in the low quarter on similarity show increased realism of perception while those in the high quarter on similarity to the adolescent negative image, show a loss of realism of perception.

Change in consistency of beliefs and attitudes and in attitude as a function of the similarity of the subject's beliefs to the commonly held adolescent negative image of the scientist proves in analysis to be a reversal of the so-far typical pattern. Those subjects in the high quarter on similarity show little change but those in the low quarter on similarity show a marked decrease in consistency and an increased negative attitude.

Change in reality perception of scientists analyzed by the degree of uncertainty ("don't know" responses on Part II-B Beliefs) expressed at the outset reveals highly significant differences--those subjects originally high in uncertainty show a gain in reality perception while those originally low in uncertainty show a marked decrease in reality perception. Analysis of change in consistency and attitude as a function of degree of original uncertainty also indicates some real differences--although both groups show a loss, the loss is significantly greater for the subjects with initial low uncertainty than for the subjects with initial high uncertainty.

There was no significance in the differences in reality perception between subjects with high and low agreement with stable peer values; however, with regard to consistency of beliefs and values and attitude toward scientists, the subjects in low agreement showed only slight change while the subjects high in agreement with peer values showed a significant change in the negative direction--inconsistency of their values with their beliefs about scientists and a negative attitude toward scientists--the direction which appears to prevail for the peer group as a whole.

By means of multiple regression analysis, the degree to which change in reality perception from first to third testing period would have to be attributed to chance factors or to sampling variation rather than to the variables being used to predict change was found to be only one time in a hundred with the value obtained for multiple correlation. Thus it can be said that these variables can be used with confidence to predict change in reality perception. Likewise, by a multiple regression analysis, the degree to which changes in consistency of beliefs about scientists and personal values and in attitude from first to third testing period would have to be attributed to chance factors or to sampling variation rather than to the variables being used to predict change was found to be only one time in a hundred. Here it can also be assumed that these variables can be used with confidence to predict change in consistency and in attitude.

Changes in reality perception appear to be a function of the following interrelationships:

Gains	Losses
1 low similarity to adult stereotype	1 high similarity to adult stereotype
2 low similarity to adolescent positive image	2 high similarity to adolescent positive image
3 low similarity to adolescent negative image	3 high similarity to adolescent negative image
4 high uncertainty of beliefs	4 low uncertainty of beliefs

Changes in consistency of beliefs with values and in attitude appear to be a function of the following interrelationships: (Change in this area took the form of losses rather than gains for the most part.)

Gains	Losses
1 (low similarity to adult stereotype)	1 high similarity to adult stereotype
2 (low similarity to adult values)	2 high similarity to adult values
3 low similarity to adolescent positive image *	3 high similarity to adolescent positive image
4 (high uncertainty of beliefs)	4 low similarity to adolescent negative image **
5 (low similarity to stable child values)	5 low uncertainty of beliefs
	6 high similarity to stable child values

* true gain; ( ) inferred gain
** reversal of established pattern

V. SUMMARY AND CONCLUSIONS

During 1959, Boston University School of Education with funds provided by the U. S. Office of Education under Title VII of the National Defense Education Act of 1958 carried out a research project to study the effects of televised instruction with 90 fifth grade classes participating in The 21" Classroom of the Eastern Massachusetts Council for School Television. The Massachusetts Audubon Society with the support of the Children's Museum and with funds supplied under the USOE grant produced the series of natural science programs called "Exploring Nature." These programs were telecast by The 21" Classroom using the facilities of Boston's educational channel, WGBH-TV.

Our objectives were to study how much children learned from these programs, what effect the series had on children's interests, attitudes, and reasoning ability, and which of certain combinations of auxiliary classroom procedures proved more effective. We attempted to obtain answers to several specific questions in relation to the behavioral changes anticipated. What effects occurred as a result of:

1. Teachers being trained in science for the elementary school as compared with training in utilization of television?
2. Teachers doing their work with classes ahead of the program rather than after?
3. Teachers having a comprehensive study guide to use compared with a brief outline of the programs?
4. Pupils being encouraged to develop individual interests and specialties instead of working on common class assignments?
5. Pupil attitudes toward science and scientists being typical of those prevailing among children and adults as opposed to being atypical?

In addition, we were concerned with the usual questions regarding the relationship of change to differences in sex, intelligence and parental occupation. The primary unit of study was the class rather than the individual pupil.

Superintendents of school systems participating in The 21" Classroom were asked to indicate which and how many classes in their system could be used in the study. Some 300 classes were offered from which were randomly drawn 90 classes--sixty from towns, 30 from cities. Eighteen of the 90 classes were designated as control classes--they would view the programs but have at most a minimal follow-up in the classroom. The remaining 72 classes constituted the experimental group and of the 72 teachers of these classes, 24 received training in a workshop in science for the elementary school, 24 participated in a workshop in the utilization of television. The remaining 24 teachers in the experimental group received no special training.

The 72 classes in the experimental group were randomly assigned to the following groups with equal numbers of teachers in the teacher-training categories in each:

1. In the fall before any of the programs were shown, half the teachers started their classes working on projects and material to be covered in the first seven programs dealing with animals. These classes were well under way on their work before seeing any programs and the programs served to terminate the unit of study. The other half of the classes waited until they had viewed the first three or four programs before starting any follow-up activities. In this case, the television programs initiated the learning and presumably served to stimulate interest and ideas. These schedules were maintained throughout the school year.

2. Within each of these two divisions--initiating and terminating use of television--half of the teachers had available the extensive study guide prepared by the Massachusetts Audubon Society to guide classroom learning in conjunction with the series; the other half of the teachers had only a brief outline of the programs to be presented and were expected to develop their own lesson plans.

3. Within each of these two divisions--study guide and no study guide--half of the teachers encouraged their pupils to work on individual projects of personal interest; the other half gave common assignments to all members of the class.

The experimental assignments were designed so that not only individual effects but also combinations of the various procedures being utilized could be observed. Each teacher reported weekly on the work being done in the class and the classes were visited periodically by a research staff member. Tests of science achievement, interests, attitudes and reasoning ability were administered at the beginning of the school year; achievement and interest tests were readministered at midyear to the experimental classes; and all tests were administered to all classes again at the end of the school year. Test results were recorded and analyzed by IBM computer.

Results are reported primarily in terms of changes, i.e., gains made on achievement tests, shifts in interest, changes in attitude. Two major sets of comparisons were made, the first between the control and the experimental groups, the second within and between experimental assignments. The major effects observed can be summarized as follows:

1. Large, statistically significant differences were obtained between the control and experimental groups in science achievement as measured by information and vocabulary tests, the difference in favor of the experimental classes. Gains in information were related to intelligence level.

2. Differences between experimental and control groups on the tests of scientific reasoning ability did not reach the level of statistical significance. Probably one of the most important discoveries, however, in the light of popular belief, is that there was no difference by sex in scientific reasoning ability. Girls did equally as well as boys. This suggests that where differences are observed (primarily among adults), if in fact they are true

differences, they may be attributable to aculturation rather than to sex-linked characteristics.

3. On the measure of interest in science, the control group showed a slight increase, the girls in the experimental group showed a slight decrease in interest and the boys remained about the same. Losses in interest were inversely related to intelligence level, the higher I. Q. quartiles losing interest, the lower quartiles gaining. Reported participation by subjects in science-related leisure activities and TV viewing followed their interest in science in general.

4. None of the main experimental effects produced statistically significant differences between classes within the experimental group. In combination, however, some of the experimental conditions did make significant differences, notably combinations of teacher training and pupil assignment, and the point at which television was used and pupil assignment.

5. On the whole children grew more realistic in their perception of science and scientists, but also grew more negative in attitude in this area. Changes were not related to the experimental conditions but were related to the similarity of subject's beliefs and attitudes to those of socially relevant child and adult groups.

Although the results as reported are factual, their interpretation must of necessity be speculative. Nevertheless, *on the basis of the results obtained, the following conclusions* would seem to be reasonable:

1. Teacher training in a specific subject would have to be more intensive than that provided in workshops established for this study to be translated into significant differences in pupil achievement. The training provided probably served more to orient the teacher to the course of study and the educational objectives and to make her more efficient in her classroom performance than to provide her with sufficient knowledge about science or audio-visual techniques to make for material differences in the learning of children in the various classrooms.

2. If television programs (and this would probably apply to other audio-visual techniques) are to be used as a terminal activity, greater learning occurs where the class assignments are made by the teacher and are common to all. In contrast, where television is used to initiate a program, and this applies whether the programs are used as a block or singly, the optimum combination to be used in follow-up activity is individual projects or assignments. It is probable that lacking guide lines, common class assignments provide clues as to appropriate direction for work. Once having seen programs, children are likely to choose topics or work consistent with what they have seen, rather than ranging afield.

3. Teachers with a brief outline and suggestions for pupil projects achieved as much with their classes as teachers with a comprehensive study guide. Whether the explanation is that teachers only utilize a limited amount of information available or that the study guides provide general background information which does not transfer readily into pupil gains is not clear, but it does suggest that the kinds of teaching materials and training materials have to be carefully appraised for their intrinsic worth as far as instructional purposes are concerned.

4. Pupil attitudes towards science and towards scientists do not change readily. It was interesting to note that the children did not perceive the television instructor as a scientist, but rather as a teacher, nor were other members of the research team conceived of as scientists. Few of these children had any first-hand contact with scientists. The resistance of their attitudes to change is consistent with other mass media research which indicates that stereotypes established in the absence of reality experience are difficult to change.

5. The most startling results are the effects on interest in science which show children of high initial interest and children of high intelligence tending to lose interest as the school year progresses, and that the more intensive the school work, the greater the loss. One consideration must be kept in mind in appraising this finding, namely that we are dealing with group effects, not individual effects. For instance, for a mean score for a class to hold constant, every loss of interest by one child has to be offset by a gain in interest by another child. Thus individual children were observed both losing and gaining in interest. Nevertheless for the classes as a whole, the findings are as stated. At least two explanations are possible: one is that the initial high interest was not based on reality and that direct experience produced a more realistic experience measure in most children; the other explanation is that for children who possess an established interest, making it a phase of school work tends to kill it. In other words intrinsic interests are pursued for the direct satisfaction they provide. Making them part of school work robs them of their spontaneity. Retesting in the fall of the coming year should help to answer the question, for spontaneous interests should tend to regain their former status over the summer, whereas unrealistic interests once lost as a result of experience should not regain their former status in the absence of further satisfying experience.

At present, the finding would suggest that for pupils of high intelligence who have high science interest, a preferable approach in the classroom would be to encourage them to pursue their interests along their own lines rather than following the path of regular assignments. There is a risk, probably slight, that these children might not learn as much factual information. More likely, they would learn more but different information than other members of the class. Yet given a choice between interest and information, the decision would best be made in favor of the former, for in the absence of motivation, information gains are likely to be transitory at best.

One fact is certain, namely that our pedagogical skill in influencing interest and attitude and teaching science reasoning ability does not approximate our skill in transmitting knowledge.

6. The considerably greater range of achievement gains in the experimental group than in the control group indicates that variations in skill among teachers yield considerable differences in learning. The low correlation between gains and amount of time spent in follow-up activity suggests that the quality of teaching is primary rather than amount of time given to science. A passive introduction of the television programs with desultory follow-up will not yield the learning that any combination of the arrangements attempted will. The task cannot be relegated to the television teacher. Rather, if The 21" Classroom is to be incorporated into a 30 foot classroom, the results will depend on the skill of the teacher in utilizing the television programs and coordinating her efforts with them.

7. Of concern is the fact that while major changes are observed in gains in factual information, the observed effects on interest, attitude, and science reasoning are negligible. The question of the validity of the instruments used to measure these aspects of behavior may be raised. The findings relative to interest were corroborated by checks of leisure activities and television viewing. The findings on attitude were consistent with our hypotheses and previous research by other investigators. The tests of science reasoning included a standardized instrument as well as the instruments constructed for purposes of this study. Possibly the instruments though valid were not sufficiently sensitive to the changes that occurred. Nevertheless, if the findings are accurate, then it suggests that while we may be effective in providing for conventional academic achievement, we have much to learn about creating an educational climate which provides for specific changes in interest, attitude and science reasoning.

BIBLIOGRAPHY

1. Adorno, T. W., E. Frenkel-Brunswik, D. J. Levinson, and R. N. Sanford, *The Authoritarian Personality*. New York: Harper and Brothers, 1950.

2. Ainsworth, L. H. "Rigidity, Insecurity and Stress," *J. Abnormal and Social Psychology*, 1958, *56*, 67-74.

3. Baggaley, Andrew R. "Concept Formation and its Relation to Cognitive Variables." *J. General Psychology*, 1955, *52*, 297-306.

4. Battig, W. F. "Some Factors Affecting Performance on a Word Formation Problem." *J. Experimental Psychology*, 1957, *54*, 96-104.

5. Bayles, Ernest E. "Is Modern Science Inductive?" *The Science Teacher*, 1954, *21*, 113-16, 145.

6. Billings, M. L. "Problem Solving in Different Fields of Endeavor." *American J. Psychology*, 1934, *46*, 259-272.

7. Bills, Robert E., in *Instruction in Learning*: Implications for Television, Report of N. E. A. Seminar, Division of Audio-Visual Instructional Service, Finette P. Foshay (ed.), January 31-February 3, 1959

8. Bloom, Benjamin S. (ed.), *Taxonomy of Educational Objectives Handbook I: Cognitive Domain*. New York: Longmans, Green and Co., 1956.

9. Brandwein, Paul F. "The College Board's Science Tests." *The Science Teacher*, 1952, *19*, 107-13.

10. Brooker, F. E. "How Television is Progressing in Schools." *School Executive*, 1950 (April), *69*, 67-69.

11. Brownell, W. A. and G. Hendrickson. "How Children Learn Information Concepts and Generalizations." *49th Yearbook National Society for the Study of Education*, 1950, Part I, 92-128.

12. Buehring, L. E. and W. M. Brish. "Hagerstown Experiment After Two Years." *Nation's Schools*, 1959 (February), *63*, 83-86.

13. Bureck, B. "The Nature and Efficiency of Methods of Attack on Reasoning Problems." *Psychological Monographs*, 1950, *64*, No. 7.

14. Burmester, Mary Alice. "Behavior Involved in Critical Aspects of Scientific Thinking." *Science Education*, 1952 (December), *36*, 259-63.

15. ______________. "The Construction and Validation of a Test to Measure Some of the Inductive Aspects of Scientific Thinking." Science Education, 1953, 37, 131-40.

16. Campbell, Marjorie H. "Teaching Science by Television." Elementary School Principal, 1953 (September), 33, 204-10.

17. Carpenter, C. R. and L. P. Greenhill, An Investigation of Closed-Circuit Television for Teaching University Courses. Instructional Research Project No. 1, Pennsylvania University, University Park, Pa., July, 1955

18. Carpenter, C. R. and L. P. Greenhill, An Investigation of Closed-Circuit Television for Teaching University Courses. Report No. 2, Pennsylvania University, University Park, Pa., Spring 1958.

19. Cartwright, Dorwin. "Social Psychology and Group Processes." Annual Review of Psychology, 1957, 8, 211-13.

20. Clark, C. C. "Television in Education." School and Society, 1938, 48, 431-2.

21. Craig, R. C. "Directed versus Independent Discovery of Established Relations." J. Educational Psychology, 1956, 47, 223-234.

22. Crowell, Victor L., Jr. "The Scientific Method, Attitudes and Skills Essential to the Scientific Method and Their Treatment in General Science and Elementary Biology Textbooks." School Science and Mathematics, 1937, 37, 525-31.

23. Croxton, W. C. "Pupils' Ability to Generalize." School Science and Mathematics, June, 36, 634.

24. Curti, Margaret W. "Child Development, X Concepts." Encyclopedia of Educational Research, New York: Macmillan Co., 1950, 175-7.

25. Desiderato, Otello L., Joseph H. Kanner, and Richard P. Runyon. "Procedures for Improving Television Instruction." Audio-Visual Communication Review, 1956 (Winter). 4, 57-63.

26. Detambel, Marvin H. and L. M. Stolurow. "Stimulus Sequence and Concept Learning." J. Experimental Psychology, 1956 51, 34-40.

27. Deutsche, Jean M., The Development of Children's Concepts of Causal Relations. Minneapolis: University of Minnesota Press, 1937.

28. Dietmeier, Homer J., *The Effect of the Integration of Science Teaching by Television on the Development of Scientific Reasoning in the Fifth Grade Student*. Unpublished Doctoral Dissertation, Boston University, 1961.

29. Draper, Benjamin. "Science Education Through Television." *National Association of Secondary School Principals Bulletin*, 1953, *37*, 149-155.

30. Dunham, Franklin and Ronald Lowdermilk. *Television in Our Schools*. Bulletin 1952 (rev. 1956), No. 16, United States Office of Education, Washington, D. C.

31. Dunning, Gordon M. "The Construction and Validation of a Test to Measure Certain Aspects of Scientific Thinking in the Area of First Year College Physics." *Science Education*, 1949, *33*, 221-235.

32. ______, "Using the Laboratory to Develop Critical Thinking." *Science Teacher*, 1952, *19*, 85.

33. Edwards, T. Bentley. "Measurement of Some Aspects of Critical Thinking." *J. Experimental Education*, 1950, *18*, 263-78.

34. Engelhart, Max D. and Hugh B. Lewis. "An Attempt to Measure Scientific Thinking." *Educational and Psychological Measurement*, 1941, *1*, 289-94.

35. Farrell, J. V. and J. R. Wailes. "Multisensory Approach to Science in the Elementary School." *Elementary School Journal*, 1952, *52*, 271-76

36. Gable, Martha A. "Television in the Classroom." *School Executive*, 1950, *69*, 70-2.

37. Garber, Robert R. "Influence of Cognitive and Affective Factors in Learning and Retaining Attitudinal Materials." *J. Abnormal and Social Psychology*, 1955, *51*, 384-89.

38. Goodnow, J. J. and T. F. Pettigrew. "Effect of Prior Patterns of Experience Upon Strategies and Learning Sets." *J. Experimental Psychology*, 1955, *49*, 381-89.

39. Gordon, L. V. and M. A. Durea. "The Effect of Discouragement on the Revised Stanford-Binet Scale." *J. Genetic Psychology*, 1948, *73*, 210-17.

40. Gross, Calvin E. "The Superintendent's Viewpoint on Educational Television." Panel Discussion, Region I Conference, N. A. E. B. Published by Thomas Alva Edison Foundation, New York, March, 1959

41. Hanfmann, E. and J. Kasanin. "A Method for the Study of Concept Formation." *Journal of Psychology*, 1937, *3*, 521-40.

42. Heidbreder, E. "The Attainment of Concepts: I. Terminology and Methodology." J. General Psychology, 1946, 35, 173-89.

43. Herminghaus, Earl, An Investigation of Television Teaching. Mimeographed, St. Louis Public Schools, St. Louis, Missouri, September, 1956, and February, 1957.

44. Hill, Katherine E. "Children's Contributions in Science Discussions." Teachers College Columbia University Contributions to Education, N 931. New York: Bureau of Publications, Teachers College Columbia University, 1947.

45. Hoban, Charles F. "Hope and Fulfillment in ETV Research." Audio-Visual Communication Review, 1958, 6, 165-71.

46. Hunter, I. M. L. "The Solving of Three-Term Series Problems." British J. Psychology, 1957, 48, 286-98.

47. Inhelder, Barbel and Jean Piaget, The Growth of Logical Thinking from Childhood to Adolescence. New York: Basic Books, 1958.

48. John, E. R. and J. G. Miller. "The Acquisition and Application of Information in the Problem Solving Process." Behavioral Sciences, 1957, 2, 291-300.

49. Kanner, J. H. "Teaching by Television in the Army: An Overview." Audio-Visual Communication Review, 1958, 6, 172-88.

50. ____________, Sanford Katz, William Mindak and Peter Goldsmith. "Television in Army Training: Evaluation of Television for Intensive Training and for Reducing Instructor and Student Training Time and Costs." Audio-Visual Communication Review, 1958, 6, 255-291.

51. ____________, Richard P. Runyon, and Otello Desiderato, Television in Army Training: Evaluation of Television in Army Basic Training. Technical Report 14. Human Resources Research Office, George Washington University, Washington, D. C. 1954.

52. Keeslar, Oreon. "A Survey of Research Studies Dealing with the Elements of Scientific Method as Objectives of Instruction in Science." Science Education, 1945, 29, 212-16.

53. Klapper, Hope L. "Prevalent Interests and Concerns in the Field of Televised Instruction." J. Educational Sociology, 1959, 32, 437-51.

54. Kraft, Mary, A Study of Information and Vocabulary Achievement in the Teaching of Natural Science by Television in the Fifth Grade. Unpublished Doctoral Dissertation, Boston University, 1961.

55. Kurtz, K. H. and C. I. Hovland. "Concept Learning with Differing Sequences of Instances." *J. Experimental Psychology*, 1956, *51*, 239-43.

56. Leeper, R. W. "Cognitive Process." *Handbook of Experimental Psychology*. S. S. Stevens (ed.) New York: Wiley and Sons, Inc., 1951, 730-57.

57. Levenson, W. B. "TV in the Classroom." *NEA Journal*, 1953, *43*, 562.

58. Levitt, E. E. "The Water-Jar Einstellung Test as a Measure of Rigidity." *Psychological Bulletin*, 1956, *53*, 347-70.

59. Likert, R. "A Technique for the Measurement of Attitudes." *Archives of Psychology*, 1932, No. 140.

60. Luchins, A. S. "On Recent Usage of the Einstellung Effect as a Test of Rigidity." *J. Consulting Psychology*, 1951, *15*, 89-94.

61. Maag, C. H. "Development and Evaluation of a Conceptual Reasoning Test." *Educational and Psychological Measurements*, 1957, *17*, 230-39.

62. Maltzman, I., E. Eisman, and L. O. Brooks. "Some Relationships Between Methods of Instruction, Personality Variables and Problem Solving Behavior." *J. Educational Psychology*, 1956, *47*, 71-8.

63. Marks, M. R. "Problem Solving as a Function of the Situation." *J. Experimental Psychology*, 1951, *41*, 74-80.

64. Mason, John M. "An Experimental Study in the Teaching of Scientific Thinking in Biological Science at the College Level." *Science Education*, 1952, *36*, 270-84.

65. McNemar, O. W. "An Attempt to Differentiate Between Individuals with High and Low Reasoning Ability." *"American J. Psychology*, 1955, *68*, 20-36.

66. McPhee, W. N. "New Strategies for Research on the Mass Media." Unpublished report from the Bureau of Applied Social Research, Columbia University, 1953 (used by permission of the author).

67. Mead, M. and R. Metraux. "Image of the Scientist Among High School Students: A Pilot Study." *Science*, 1957, *126*, 384-90.

68. Mereško, R., M. Rubin, F. C. Shontz and W. Morrow. "Rigidity of Attitude Regarding Personal Habits and Its Ideological Correlates." *J. Abnormal and Social Psychology*, 1954, *49*, 378-80.

69. Murphy, G. *Human Potentialities*. New York: Basic Books, Inc., 1958, 340 pp.

70. Nelson, Pearl A. *The Acquisition of Concepts of Light and Sound in the Intermediate Grades*. Unpublished Doctoral Dissertation, Boston University, 1957.

71. Newman, S. E. "Effects of Contiguity and Similarity on the Learning of Concepts." *J. Experimental Psychology*, 1956, *52*, 349-53.

72. O'Connell, S. C. H. Ernestine, Maria, Sister. *The Comparison of Inductive and Deductive Methods of Teaching High School Chemistry*. Unpublished Doctoral Dissertation, Boston University, 1958.

73. Peak, H. and H. W. Morrison, "Attitude Structure," *J. Abnormal and Social Psychology*, 1958, *57*, 127-35.

74. Reed, H. B. "Factors Influencing the Learning and Retention of Concepts: I. The Influence of Set." *J. Experimental Psychology*, 1946, *36*, 71-87.

75. Reiner, William B. "The Effectiveness of a Television Series in Improving Kindergarten to Grade II Science Teaching Programs." *School Science and Mathematics*, 1959, *59*, 397-408.

76. Remmers, H. H. *Introduction to Opinion and Attitude Measurement*. New York: Harper and Bros., 1954.

77. Remmers, H. H. and D. H. Radler. *The American Teenager*. Indianapolis-New York: Bobbs-Merrill Co., 1957.

78. Rokeach, M. "Generalized Mental Rigidity as a Factor in Ethnocentrism." *J. Abnormal and Social Psychology*, 1948, *43*, 259-78.

79. Rulon, Philip J. "The Sound Motion Picture in Science Teaching." *Harvard Studies in Education*, Cambridge, Mass.: Harvard University Press, 1933, Vol. 20, 236 pp.

80. Saugstad, P. "Incidental Memory and Problem Solving." *Psychological Review*, 1952, *59*, 221-26.

81. Sheehan, A. Cornelia. *The Interrelationships of Interests and Attitudes and Specified Independent Variables in the Teaching of Natural Science by Television in the Fifth Grade*. Unpublished Doctoral Dissertation, Boston University, 1960.

82. Siepmann, Charles A., *National Program in the Use of Television in the Public Schools: Some Reflections on the First Year*. Fund for the Advancement of Education (mimeographed), New York, June 10, 1958.

83. Simpson, G. G., C. S. Pittendrigh and L. H. Tiffany. *Life: An Introduction to Biology*. New York: Harcourt, Brace and Co., 1957.

84. Steele, Ralph. "Educational Television--1958!" *Educational Screen*, 1958, *37*, 614-15.

85. *Television in Instruction: An Appraisal*. Department of Audio-Visual Instruction, National Education Association, Washington, D. C., 1958.

86. Tyler, K. (ed.) *NAEB Research Seminar for Educational Television and Radio*. Columbus: Ohio State University, 1957.

87. Underwood, B. J. and J. Richardson. "Verbal Concept Learning as a Function of Instructions and Dominance Level." *J. Experimental Psychology*, 1956, *51*, 229-38.

88. Vinacke, W. E. *The Psychology of Thinking*. New York: McGraw-Hill Book Co., 1952

89. Webb, Ruth. "All Children Think and Plan." *Childhood Education*, 1947, *23*, 315-21.

90. West, J. Y. "Do We Expect Too Much or Too Little of Children from Their Experiences in Science?" *Science Education*, 1949, *33*, 296-98.

91. Wigren, Harold. "TV: Power Tool for Teachers." *NEA Journal*, 1959, *48*, 28-9.

92. Woodburn, John H. "Visual Aids and Problem-Solving Type of Teaching Exercise." *Science Teacher*, 1953, *20*, 167-9.

93. Yanowski, Barbara. "Teaching Science by Television." *Science Teacher*, 1958, *25*, 406-7.

94. Youtz, R. P. "The Relation Between Number of Confirmations of One Hypothesis and the Speed of Accepting a New and Incompatible Hypothesis." *American Psychologist*, 1948, *3*, 248-9.